ISSN 2579-2784 (Print)
ISSN 2538-2788 (Online)

MATHEMATICAL PROBLEMS OF COMPUTER SCIENCE

LVI

Yerevan
2021

Հայաստանի Հանրապետության Գիտությունների ազգային ակադեմիայի
Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ

Институт проблем информатики и автоматизации Национальной академии наук Республики Армения

Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia

Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ

Математические проблемы компьютерных наук

Mathematical Problems of Computer Science

LVI

ՀՐԱՏԱՐԱԿՎԱԾ Է ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏԻ ԿՈՂՄԻՑ
ОПУБЛИКОВАНО ИНСТИТУТОМ ПРОБЛЕМ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ НАН РА
PUBLISHED BY INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS OF NAS RA

ԵՐԵՎԱՆ 2021 YEREVAN

Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ, LVI

Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ պարբերականը հրատարակվում է տարեկան երկու անգամ ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի (ԻԱՊԻ) կողմից: Այն ընդգրկում է տեսական և կիրառական մաթեմատիկայի, ինֆորմատիկայի և հաշվողական տեխնիկայի ժամանակակից ուղղությունները:

Այն ընդգրկված է Բարձրագույն որակավորման հանձնաժողովի ընդունելի ամսագրերի ցանկում:

Տպագրվում է ՀՀ ԳԱԱ ԻԱՊԻ Գիտական խորհրդի 2021թ. դեկտեմբերի 14-ի N 21-12/1 նիստի որոշման հիման վրա

ISSN 2579-2784(Print)

ISSN 2538-2788(Online)

Математические проблемы компьютерных наук, LVI

Журнал **Математические проблемы компьютерных наук** издается два раза в год Институтом проблем информатики и автоматизации НАН РА. Он охватывает современные направления теоретической и прикладной математики, информатики и вычислительной техники.

Он включен в список допустимых журналов Высшей квалификационной комиссии.

Печатается на основании решения N 21-12/1 заседания Ученого совета ИПИА НАН РА от 14 декабря 2021г.

ISSN 2579-2784(Print)

ISSN 2538-2788(Online

Mathematical Problems of Computer Science, LVI

The periodical **Mathematical Problems of Computer Science** is published twice per year by the Institute for Informatics and Automation Problems of NAS RA. It covers modern directions of theoretical and applied mathematics, informatics and computer science.

It is included in the list of acceptable journals of the Higher Qualification Committee.

Printed on the basis of decision N 21-12/1 of session of the Scientific Council of IIAP NAS RA dated December 14, 2021.

ISSN 2579-2784(Print)

ISSN 2538-2788(Online)

CONTENTS

Mathematical Problems of Computer Science 56, 7–17, 2021.
doi: 10.51408/1963-0075

UDC **004.052**

Functional Safety Compliant Test & Repair Framework for System-on-Chip Lifecycle Management

Gurgen E. Harutyunyan, Samvel K. Shoukourian, Grigor A. Tshagharyan and Yervant A. Zorian

Synopsys, Armenia
e-mail: gurgen.harutyunyan@synopsys.com, samvel.shoukourian@ synopsys.com,
grigor.tshagharyan@synopsys.com, yervant.zorian@synopsys.com

Abstract

The share of safety-critical systems in electronic and electrical (E/E) devices across multiple domains, especially in automotive industry, is growing at a constant rate. An unhandled failure of any component within the system may compromise the safety of the entire ecosystem. Therefore, regardless of the context of use and level of the hierarchy, all system components must follow the requirements of appropriate safety standard for the development process and in-field operation. In this context, the traditional built-in self-test (BIST) scheme must also meet the safety requirements defined in ISO 26262 in order to be approved for automotive applications. The paper presents a functional safety compliant BIST infrastructure concept that helps to ensure safe test execution throughout the entire System-on-Chip (SoC) lifecycle while maintaining high test quality.
Keywords: Functional safety, Automotive, ISO 26262, ASIL, Built-in self-test, in-field test, Test and repair.
Article info: Received 15 June 2021; accepted October 2021.

1. Introduction

The automotive industry is currently one of the fastest growing sectors in the semiconductor industry. The growing demands of consumers for safety, reliability and increased safety requirements continue to drive the automotive market growth. The trend towards greater safety and better driving experience is forcing automakers to continually integrate a large number of electronic and electrical components into their vehicles, such as advanced driver assistance systems (ADAS) and in-vehicle infotainment (IVI). A few examples of such systems are adaptive cruise control, parking assist, emergency vehicle braking, lane change assist, etc. as the list continues to grow.

Functional safety sensitive systems are traditionally prone to using well-established technologies that provide high performance, reliability and lower production cost over the system lifecycle. However, given the growing demand, automakers have had to adopt solutions with increased processing power and communication performance, which require usage of relatively new technology nodes such as 3D transistors. Especially during the recent years, with increasing pressure and high competition in the consumer market, the time to market margin has drastically decreased. This trend creates additional challenges for automotive application developers and original equipment manufacturers (OEMs) to meet the demands for higher performance and energy efficiency along with natural requirements for safety, reliability and quality. As a result, the modern automotive industry faces a number of challenges that need to be addressed, including functional safety and reliability, quality of testing in the production process, as well as security and robustness. These various requirements can sometimes become conflicting and even mutually exclusive, making them difficult to fully meet.

One of the most important aspects to consider in the automotive industry is the requirement for high quality testing. Here, special attention should be paid to the development of a reliable solution for test and repair, applicable not only at the production stage, but throughout the entire life cycle of the product, from the design stage to silicon bring up and series production right up to in-field operation. This is especially related to embedded memories, covering most of the system-on-a-chip (SoC) area and making a major contribution to achieving high performance and defective parts per billion (DPPB) rate. BIST has traditionally been the preferred approach for testing and repairing embedded memories, providing a reasonable trade-off between cost and achievable yield. Until recently, the existing BIST solutions were entirely focused on the SoC production phase to provide high fault coverage and optimal yield. However, today the situation has changed dramatically, especially in the automotive industry, given the recent breakthrough in the market. Currently, safety and security in mission mode are considered the highest priority requirements for vehicles, therefore, comprehensive testing capabilities are required not only at the production stage, but also in the field.

Many techniques and new methods have already been proposed in the literature to address post-production testing problems that address specific aspects of testing. In [1] and [2], the concept of transparent memory BIST is described, where the key idea is to make the test execution transparent to the system, keeping the contents of the memory intact. Other versions of the transparent BIST architecture with improved coverage and test times were also introduced later in [3]-[5]. The implementation of memory and logic BIST for automotive SoCs described in [6] introduces several new features specifically designed for in-field testing. For example, the idea of non-destructive and destructive self-tests is presented. In the former case, test control units are not included in the test, allowing for more control in the field and planning of test procedures. Meanwhile, in the latter case, the entire device is tested, and after the test, a hard reset or system restart is required, as the system state is lost. In [7], a short burst-based BIST technique is introduced, where the idea is to divide the memory into smaller chunks and partition the BIST execution into a series of short bursts testing one chunk of the memory at a time.

There are other works as well on this topic showing the benefits of using structural tests (mostly built-in) for field testing and diagnosis purposes. For example, [8] shows the trade-offs and benefits that the automotive field can gain by reusing production testing methods for in-system testing. In addition to structural testing, other alternative methods recently proposed in the literature can also be adopted for in-field testing. This includes software-based self-test methods described in [9]-[10] and functional test methods, a good review of which can be found in [11].

Two other aspects related to in-field testing are the problem of integrated circuit aging in

mission mode and the importance of lifetime testing. With regard to aging, several methods of online aging monitoring are proposed, such as the architecture presented in [12], which is specially designed for safety-critical applications. The proposed built-in sensors work only when the car is turned on, therefore they are resistant to aging and threshold voltage fluctuations. Aging monitoring is done by observing propagation delays in critical parts of the chain. In [13], the authors carried out a study on aging faults investigation in the field and proposed an algorithmic method for the detection of aging-induced faults in an earlier stage of SoC lifetime. Meanwhile, the importance of adopting a functional approach to characterize the reliability and operating lifetime of SoCs is discussed in [14]. A new approach to automatically generate appropriate test patterns for use in the mission mode is presented, with experimental results demonstrating the advantages of the proposed method.

The purpose of this paper is to combine the existing best practices and present the concept of a full-scale functional safety compliant solution for testing embedded memories in automotive SoCs with the help of the proposed universal BIST engine. In other words, the proposed BIST concept represents a one-stop test solution from production to in-field test. The next section of this paper provides an introduction to the ISO 26262 [14] standard and the certification process in automotive. Section 3 takes a closer look at the phases of automotive in-field testing and the various requirements they impose. The details of the proposed solution for automotive SoCs are described in Section 4. Finally, Section 5 draws the conclusions.

2. Functional Safety and ISO 26262

One of the key requirements to consider when developing any application for an automotive SoC is functional safety and reliability. Known requirements for applications critical in terms of functional safety, established in industries such as the military, nuclear and aerospace, are now being widely transferred to the automotive industry. The main goal of safety and reliability is to tolerate the risk of physical injury or of damage to the health of people. Safety primarily aims to reduce the risk of systematic as well as random hardware failures in the system during manufacturing or in-field operations. At the same time, reliability determines the probability that the system will perform the functions assigned to it in a given period of time.

The increased attention to safety and reliability aspects in the automotive industry has led to the need for common criteria to measure the level of their compliance in the system. This was the reason for the emergence of ISO 26262 standard called “Road vehicles: Functional Safety”, which establishes the definitions and requirements for functional safety for automotive equipment applicable throughout the life cycle of all safety-oriented automotive E/E systems. ISO 26262 specifies requirements for achieving an acceptable level of functional safety for electrical and/or electronic systems intended for use in vehicles. Depending on meeting these requirements, the final product can be qualified with one of the four available Automotive Safety Integrity Levels (ASIL) A to D.

ASIL refers to an abstract classification of the inherent safety risk in an automotive system or elements of such a system. The ASIL classification is used in ISO 26262 to express the level of risk reduction required to avoid a particular hazard, where ASIL-D is the highest level and ASIL-A is the lowest level. The ASIL evaluated for that hazard is then assigned to the safety goal established to eliminate that hazard, and then inherited by the safety requirements resulting from that goal. ASIL is determined based on a combination of the probability of exposure, the possible controllability by the driver, and the severity of the possible consequences in case critical event occurs.

The general ASIL certification process for random hardware faults in automotive products consists of the following main steps:

1. The product is handed over to the certification body.
2. Safety Goal Violations (SGVs) of the product are determined:
   - Safety Goal (SG) is a safety requirement imposed on a product in order to reduce the risk of one or more hazardous events to an acceptable level;
   - Safety Goal Violation (SGV) is a violation of a safety goal due to product malfunction.
3. The failure modes (FM) of the product are defined:
   - Faults that lead directly to the violation of a safety goal are called Single Point Faults (SPFs);
   - MPFs (Multiple Point Faults) are a combination of two or more independent faults leading directly to the violation of a safety goal.
4. Diagnostics Coverage (DC) for each FM and SGV is calculated:
   - At this stage, the certification body determines the impact of each FM on each SGV and calculates the corresponding DC of the product able to detect that impact.
5. ASIL-X level is defined based on the DC numbers:
   - Based on the obtained DC number, the ASIL level of the product is determined from Table 1.
6. FIT rate of the product is calculated:
   - Using known formulas, the certification body calculates the FIT rate of the product.
7. FMEDA report is prepared:
   - The FMEDA report contains all the information obtained during the certification process including SGVs, FMs, DCs, ASIL and FIT numbers.
8. The certification body provides the ASIL-X level compliance certificate.

Table 1. ASIL and FIT requirements

ASIL	**SPF**	**MPF**	**FIT Rate**
ASIL B	≥ 90 %	≥ 60 %	100 (recommended)
ASIL C	≥ 97 %	≥ 80 %	100 (required)
ASIL D	≥ 99 %	≥ 90 %	10 (required)

Probabilistic Metric for random Hardware Failures (PMHF) is yet another quantitative analysis method used in the scope of ISO 26262, which defines the average probability of failure per hour over the operational lifetime of the item. The formula for PMHF estimation looks like the following:

$$PMHF = \lambda_{SPF} + \lambda_{RF} + \lambda_{DPF\_detected} \times \lambda_{DPF\_latent} \times T_{Lifetime}$$

where

λ_{SPF} is the single point failure rate;

λ_{RF} is the residual failure rate;

$\lambda_{DPF\_detected}$ is the detected and notified dual point failure rate;

λ_{DPF_latent} is the latent dual point failure rate (mitigated but not notified);

$TL_{ifetime}$ is the vehicle lifetime.

2. Silicon Lifecycle Test Requirements

ISO 26262 defines clear requirements for all products before they can be used in vehicles, and the components responsible for built-in self-test and repair are no exception. In the past, the primary requirement was the test quality and the timely detection of manufacturing defects. In the automotive industry, the picture is different since the safety enters the scene. In order to better understand automotive requirements, we need to take a closer look at the various stages of the SoC life cycle and the key test requirements that need to be considered in each stage. In fact, the SoC life cycle can be divided into three main modes: production, power-on and mission modes.

A. Production Mode

In the automotive industry, as in any other high-tech industry, achieving efficient yield is a vital requirement. Therefore, it is crucial to have a comprehensive test and repair mechanism with all the necessary capabilities. To do this, at the design stage, test structures are most often built into the chip for test, repair and diagnosis purposes. The built-in structures provide the chip ability to self-test itself, reducing the complexity of test setup and the cost of using sophisticated test equipment as well as shortening time to market.

The first step to building a comprehensive test BIST infrastructure is to understand the full range of realistic fault models for the various memory architectures and technologies that will be used on the chip. Memory faults are an abstract representation of physical defects that can occur during the chip production phase. However, not only the faults occurring in the memory array need to be taken into account, but also the faults occurring in the surrounding blocks of the memory array, including the address decoder, the write driver, the sense amplifier, etc., must also be taken into account. Once a set of target faults has been determined, test algorithms need to be developed to ensure complete fault coverage and optimal yield.

B. Power-On Mode

Maintaining test structures active in the SoC even after the production phase is a particular requirement for applications focused on functional safety. Over time, the circuits wear out and negative effects can begin to appear within a few years after production, or even a couple of months in the worst case. The aging effect is the most common problem and is usually defined as the degradation of circuit performance over time. The effects of aging can include increased power consumption, reduced speed, timing delays, which can eventually lead to failures and malfunction of the system. With the rapid reduction of technology nodes, the effects of aging have increased significantly and can no longer be ignored, especially in applications such as automotive. The main causes of aging are the effects of negative bias temperature instability (NBTI) and hot carrier injection (HCI), but recently the positive bias temperature instability (PBTI) effect has also increased in importance.

Unlike the production phase, in-field test requirements are more stringent due to space, power and time constraints. Therefore, a test solution from a production mode cannot fully meet these criteria, so alternative solutions are usually offered. It is not necessary to consider all manufacturing defects during power-on testing, since as the study [13] shows only a subset of these defects may appear due to aging or electromigration. The power-on test starts whenever the engine is turned on and the main purpose is to quickly check if all devices are working properly before the system goes into mission mode, or otherwise report if any problem is detected. For the power-on test, it appears that the only known limitation is the test time, which must remain in the

certain range. Therefore, unlike production testing, this mode uses lighter complexity test algorithms that target only the most common types of permanent faults. Sometimes, depending on the application, the power off or key off mode is also used for this purpose to perform additional checks before shutting down the engine.

C. Mission Mode

After successful completion of the power-on phase, the system enters the mission mode. Mission mode is the most critical mode, as the reaction time is extremely limited if something goes wrong, otherwise it can lead to fatal consequences. In mission mode, failures usually occur either due to soft errors, or permanent faults resulting from aging. Therefore, in this mode two categories of test mechanisms can be distinguished, depending on whether they are always active or are activated periodically.

The first category includes mechanisms such as Error Correction Codes (ECC), which constantly check whether data or address integrity is maintained during in-field operation. The main requirement in this case is to minimize the time from the detection of a problem to warning message delivery to the system. The primary goal of such test mechanisms is the detection of soft errors, which are usually caused by alpha particles and cosmic rays hitting the integrated circuits [16]-[17]. Soft errors, compared to hard errors, are temporary and do not cause permanent damage. However, with the growing safety and reliability concerns in the automotive industry, detecting and correcting these types of errors in mission mode is critical.

The second category includes test mechanisms that are periodically activated to test system components that are not always available. For example, memory and logic BIST cannot run continuously, as memory and other logic blocks may be occupied by the system. Only when the memory or the other blocks are freed by the system they can be checked for faults. The tests in this category are called mission periodic tests. The periodic check runs once during the safety interval, which is defined in the system specification as the period of time during which a failure can occur. There are a number of challenges that need to be addressed by a periodic memory test:

1. The specific memory instances can be tested only when they are in an idle state
2. The allotted time for a memory test is usually not enough to test the entire memory macro
3. The content of the tested memories must be preserved intact by the test
4. When an interrupt command is received from the system, the memory test should be stopped and restarted only after the memory is freed.

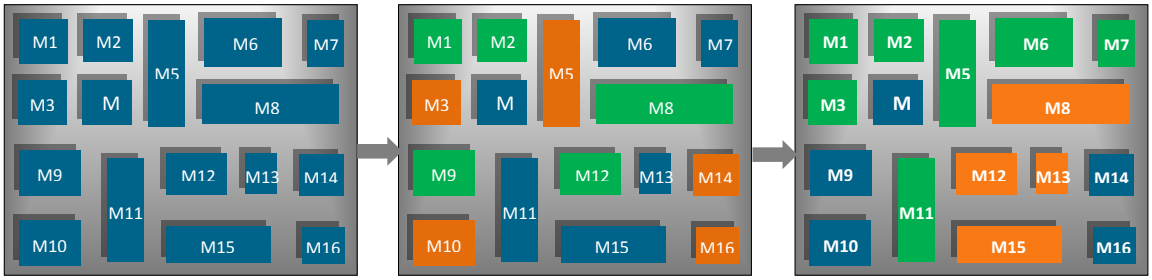

Fig. 1. Periodic test in the field: blue-colored blocks are in idle state,
green-colored blocks are in mission and orange-colored blocks are in test mode.

Therefore, advanced methods are applied to divide the test into parts, and each time execute one part of the entire test. This means only a predefined subset of memory cells is tested during each test session. It is important to note that the state after each test session must be saved in the

system in order to continue the test from exactly the same place when starting the next session. Fig. 1 shows an example of periodic in-field test at different points in time. In the first scenario, all memory blocks are idle, neither is being used by the system, nor being tested. In the second scenario, the memory blocks M1, M2, M8, M9, and M12 are in mission mode while M3, M5, M10, M14 and M16 are in test mode. Finally, in the third scenario, memory blocks M8, M12, M13 and M15 have been released by the system and can be put into test mode while M3 and M5 have been fully tested and are called into mission by the system.

2. BIST Architecture for Automotive SoCs

Fig. 2 shows a proposed BIST memory architecture for automotive SoCs. It consists of the following blocks:

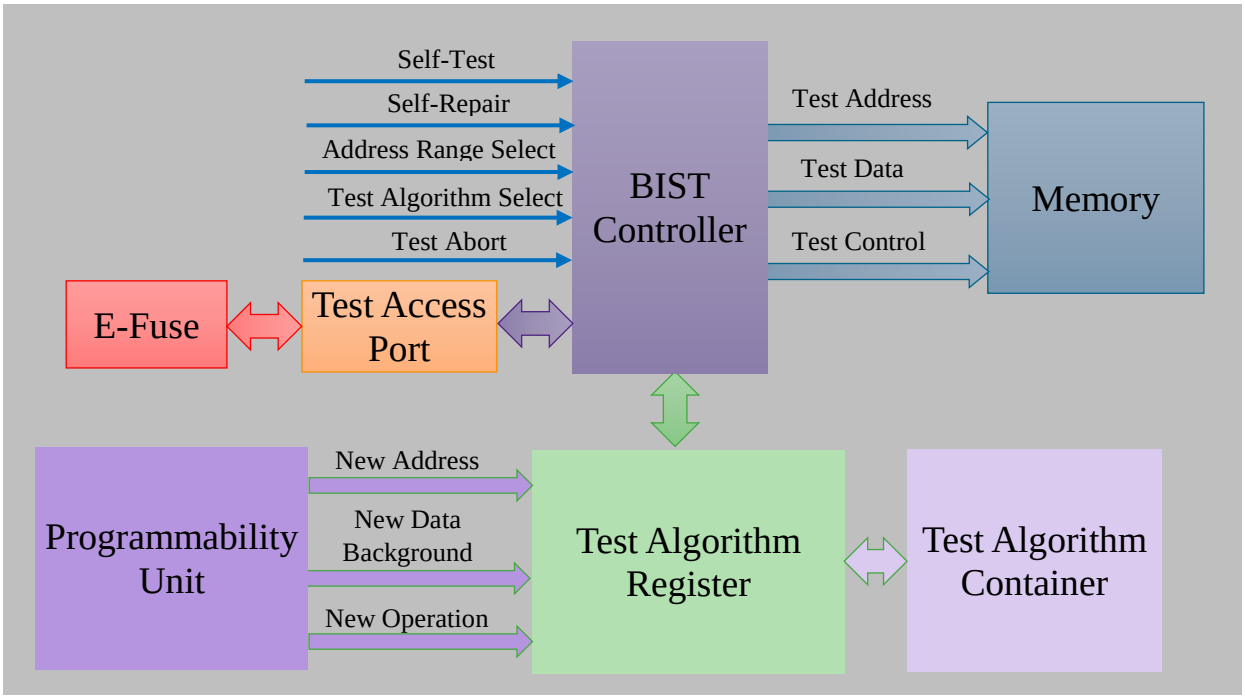

Fig. 2. The Proposed Automotive BIST Architecture.

- Test Algorithm Register - the proposed BIST architecture contains a set of predefined test algorithms designed to test and provide high fault coverage for different memory technologies. Specifically, fault coverage includes:
    - static single-cell and coupling faults
    - dynamic single-cell and coupling faults
    - links between static and dynamic faults
    - process variation faults
    - other technology-specific faults

  In addition to the comprehensive set of predefined test algorithms, it is possible to program new test algorithms or modify the existing ones if necessary. For this purpose, a programming interface is provided not only for adding new test algorithms, but also for new test operations, addressing types and background data patterns. In the case of automotive, in production mode, runtime and performance requirements are usually

sacrificed in favor of higher fault coverage, which means that complex but robust test algorithms are generally recommended.

- Test Algorithm Container - allows to store test algorithms at the design stage, and then select the appropriate one for launching in power-on and mission periodic modes. Thus, by default, along with the predefined complex test algorithms for production mode, two or more shorter test algorithms are included in the test suite with reduced length, designed for power-on and mission modes. Unlike production mode, access to the built-in test in these modes can be granted through a dedicated test interface since after production access via JTAG test access port (TAP) is usually prohibited for security reasons. In addition to the self-test capability, a quick self-repair mechanism is also provided through the same interface to quickly start the repair procedure after the test is completed.
- BIST Controller - applies selected test algorithms to memory, and also provides a number of functions necessary for testing automotive SoCs:
    - The “Self-Test” function allows to check if the BIST circuit itself is working correctly. This especially concerns the logic units, which are responsible for collecting and sending the SoC state information to the outside world. The reason for such a test is that in case anything goes wrong, incorrect system status information may be captured, compromising the overall safety of the vehicle and the driver. A flexible error injection mechanism is provided for this purpose, which allows to inject errors into the memory array and surrounding blocks.
    - The proposed solution also helps to address the challenges related to periodic test. Due to critical time constraints, complex test algorithms cannot be used for in-field test, however, periodically running simpler test algorithm compensates and minimizes the risk of possible fault escapes to some extent. Depending on the specified interval length, either the test can be performed over the entire memory space or over a range of memory addresses using the “Address Range Select” function.
    - Another more advanced feature is the capability to make the test algorithm execution transparent. This is especially important to store the content of memories containing sensitive data during the periodic test. The scheduling of a periodic test execution is determined by the system's main processor, as it depends on many factors, including the length of the available test interval, memory availability, power restrictions, etc. For that purpose, the ability to run a periodic test on selected memory blocks is also provided.
- Finally, the “Test Abort” function is provided, which allows to stop the test execution and return the memory to mission mode in case of a request from the system.

The proposed BIST architecture makes it possible not only to test, but also to repair faulty memories. During the test phase a set of detected faults with information about their location is stored in a dedicated area in the top control unit. For this purpose, an array of one-time or reprogrammable electrical fuses is usually used to store this data. After the BIST run is complete, a self-repair procedure is started based on the stored information. At the beginning of this process, a redundancy analysis is performed to read the configuration of redundant elements (rows and columns) and fault locations, and then to determine the best redundancy allocation scheme for repair. If such a scheme is found, then the repair procedure is launched and all the faulty rows and columns in the memory array are replaced with redundant counterparts. Only after the successful repair, memory is declared ready for operation.

Optionally, a powerful diagnostic capability is also part of the proposed solution in case debugging and understanding of the root cause of detected faults is required. The developed test

algorithms take as input a list of detected faults and determine the type of faults and attempt to identify common physical defect patterns, whether it is a single cell defect, a two-cell defect, a quadruple defect, a row/column defect, and so on. After the diagnosis is completed, a detailed report is generated with information about the types of faults found and the observed patterns of physical defects.

5. Conclusions

Functional safety has become increasingly important over the past decade, and especially with the introduction of the ISO 26262 standard, new requirements are placed on automotive SoCs and, ultimately, on all the components that they consist of. This paper presents the concept of an embedded memory self-test and repair solution that meets all functional safety requirements throughout the SoC's life cycle, including production, power-on, and mission modes.

References

[1] M. Nicolaidis, “Transparent bist for rams”, *International Test Conference* (ITC), pp. 598-607, 1992.

[2] M. Nicolaidis, “Theory of Transparent BIST for RAMs”, *IEEE Transactions On Computers*, vol. 45, no. 10, pp. 1141-1156, October 1996.

[3] M.G. Karpovsky, V.N. Yarmolik, “Transparent memory BIST”, *IEEE International Workshop on “Memory Technology, Design and Testing*”, pp. 106-111, 1994.

[4] S. Boutobza, M. Nicolaidis, K.M. Lamara, A. Costa, “A transparent based programmable Memory BIST”, *IEEE European Test Symposium* (ETS), pp. 89-96, 2006.

[5] I. Voyiatzis, C. Efstathiou, C. Sgouropoulou, “Symmetric transparent online BIST for arrays of word-organized RAMs”, *International Conference on Design & Technology of Integrated Systems in Nanoscale Era* (DTIS), pp. 122-127, 2013.

[6] A. Dutta, S. Alampally, A. Kumar andR. A. Parekhji, “A bist implementation framework for supporting field testability and confligurability in an automotive SOC”, Workshop on Dependable and Secure Nanocomputing, 2007.

[7] A. Becker, “Short burst software transparent on-line MBIST,” *IEEE VLSI Test Symposium* (VTS), pp. 1-6, 2016.

[8] A. Cook, D. Ull, M. Elm, H. Wunderlich, H. Randoll and S. Dohren, “Reuse of structural volume test methods for in-system testing of automotive ASICs”, *IEEE Asian Test Symposium* (ATS), pp. 214-219, 2012.

[9] F. Reimann, M. Glass, A. Cook, L. Rodriguez Gomez, J. Teich, D. Ull, H. J. Wunderlich, U. Abelein, and P. Engelke, “Advanced diagnosis: SBST and BIST integration in automotive E/E architectures,”, ACM/EDAC/IEEE Design Automation Conference (DAC), pp. 1–6, 2014.

[10] P. Bernardi, R. Cantoro, S. D. Luca, E. Sánchez, A. Sansonetti, “Development flow for on-line core self-test of automotive microcontrollers”, *IEEE Transactions on Computers*, vol. 65, no. 3, pp. 744 – 754, March 2016.

[11] A. Jutman, M. S. Reorda and H.-J. Wunderlich, “High quality system level test and diagnosis”, *IEEE Asian Test Symposium* (ATS), pp. 298-305, 2014.

[12] J. C. Vázquez et al., “Built-in aging monitoring for safety-critical applications”, *IEEE European Test Symposium* (ETS), pp. 9-14, 2009.

[13] G. Tshagharyan, G. Harutyunyan, Y. Zorian, A. Gebregiorgis, M. S. Golanbari, R. Bishnoi and M. B. Tahoori, “Modeling and testing of aging faults in FinFET memories for automotive applications”, *IEEE International Test Conference* (ITC), pp. 1-10, 2018.
[14] D. Appello et al., “Automatic functional stress pattern generation for SoC reliability characterization”, *IEEE European Test Symposium* (ETS), pp. 93-98, 2009.
[15] ISO 26262:2018 Road vehicles - Functional safety [Online]. Available: https://www.iso.org/standard/68383.html.
[16] (Jan. 2004) Tezzaron Semiconductor, “Soft Errors in Electronic Memory – A White Paper”, [Online]. Available: http://www.tezzaron.com.
[17] R. C. Baumann, “Radiation-induced soft errors in advanced semiconductor technologies”, *IEEE Transactions on Device and Materials Reliability*, vol. 5, no. 3, pp. 305-316, Sep. 2005.

Ֆունկցիոնալ անվտանգությանը համապատասխանող թեստավորման և վերանորոգման լուծում բյուրեղի վրա համակարգի կյանքի ցիկլի կառավարման համար

Գուրգեն Է. Հարությունյան, Սամվել Կ. Շուքուրյան, Գրիգոր Ա. Ճաղարյան և Երվանդ Ա. Զորյան

Սինոփսիս
e-mail: gurgen.harutyunyan@synopsys.com, samvel.shoukourian@ synopsys.com, grigor.tshagharyan@synopsys.com, yervant.zorian@synopsys.com

Ամփոփում

Անվտանգության նկատմամբ զգայուն համակարգերի մասնաբաժինը էլեկտրոնային և էլեկտրական սարքերում տարբեր ոլորտներում, հատկապես ավտոմոբիլային արդյունաբերության մեջ, աճում է շարունակական տեմպերով: Համակարգի ներսում որևէ բաղադրիչի չվերահսկվող խափանումը կարող է վտանգել ամբողջ համակարգի ապահովությունը: Հետևաբար, անկախ օգտագործման կոնտեքստից և հիերարխիայի մակարդակից, համակարգի բոլոր բաղադրիչները պետք է հետևեն մշակման և աշխատանքային ռեժիմում շահագործման գործընթացի համար անվտանգության համապատասխան ստանդարտի պահանջներին: Այս համատեքստում ավանդական ներկառուցված ինքնաթեստավորման սխեման (ՆԹՀ) պետք է նաև համապատասխանի ISO 26262-ով սահմանված անվտանգության պահանջներին, որպեսզի թույլատրվի ավտոմոբիլային կիրառությունների համար: Հոդվածը ներկայացնում է ֆունկցիոնալ անվտանգության համապատասխան ՆԹՀ ենթակառուցվածքի հայեցակարգ, որը երաշխավորում է ապահով թեստավորում

համակարգի աշխատանքի ամբողջ ցիկլի ընթացքում՝ պահպանելով թեստավորման բարձր որակը:

Բանալի բառեր՝ ֆունկցիոնալ անվտանգություն, ավտոմոբիլաշինություն, ISO 26262, ներկառուցված ինքնաթեստավորման համակարգ, թեստավորում աշխատանքային ռեժիմում, թեստավորում և վերանորոգում:

Система тестирования и ремонта, соответствующая требованиям функциональной безопасности для управления жизненным циклом системы на кристалле

Гурген Е. Арутюнян, Самвел К. Шукурян, Григор А. Джагарян и Ерванд А. Зорян

Синопсис
e-mail: gurgen.harutyunyan@synopsys.com, samvel.shoukourian@ synopsys.com, grigor.tshagharyan@synopsys.com, yervant.zorian@synopsys.com

Аннотация

Доля критических с точки зрения безопасности систем в электронных и электрических устройствах во многих областях, особенно в автомобильной промышленности, растет на постоянной основе. Необработанный отказ любого из компонентов системы может поставить под угрозу безопасность всецелой системы. Поэтому, независимо от контекста использования и уровня иерархии, все компоненты системы должны соответствовать требованиям подходяшего стандарта безопасности для процесса разработки и эксплуатации в полевом режиме. В этом контексте традиционная схема встроенной системы самотестирования (BCT) также должна соответствовать требованиям безопасности, определенным в ISO 26262, чтобы быть одобренной для автомобильных приложений. В этой работе представлена концепция инфраструктуры BCT, отвечающая требованиям функциональной безопасности, которая помогает обеспечить безопасное тестирование на протяжении всего жизненного цикла системы на кристалле, при этом сохраняя высокое качество тестирования.

Ключевые слова: функциональная безопасность, автомобильная промышленность, ISO 26262, система встроенного самотестирования, тестирование в полевом режиме, тестирование и ремонт.

Mathematical Problems of Computer Science 56, 18–34, 2021.
doi: 10.51408/1963-0076

UDC 004.932

Single Image Joint Motion Deblurring and Super-Resolution Using the Multi-Scale Channel Attention Modules

Misak T. Shoyan

National Polytechnic University of Armenia
e-mail: misakshoyan@gmail.com

Abstract

During the last decade, deep convolutional neural networks have significantly advanced the single image super-resolution techniques reconstructing realistic textural and spatial details. In classical image super-resolution problems, it is assumed that the low-resolution image has a certain downsampling degradation. However, complicated image degradations are inevitable in real-world scenarios, and motion blur is a common type of image degradation due to camera or scene motion during the image capturing process. This work proposes a fully convolutional neural network to reconstruct high-resolution sharp images from the given motion blurry low-resolution images. The deblurring subnetwork is based on multi-stage progressive architecture, while the super-resolution subnetwork is designed using the multi-scale channel attention modules. A simple and effective training strategy is employed where a pre-trained frozen deblurring module is used to train the super-resolution module. The deblurring module is unfrozen in the last training phase. Experiments show that, unlike the other methods, the proposed method reconstructs relatively small structures and textural details while successfully removing the complex motion blur. The implementation code and the pre-trained model are publicly available at https://github.com/misakshoyan/joint-motion-deblur-and-sr.
Keywords: Motion deblurring, super-resolution, channel attention.
Article info: Received 18 May 2021; accepted 14 September 2021.

1. Introduction

Single image super-resolution (SISR) addresses the problem of recovering a sharp, high-resolution (HR) image from a given low-resolution (LR) image. The super-resolution (SR) problem has become very popular during the last decade. Its solution is beneficial for a wide range of applications such as object detection, object recognition, surveillance, etc. Image SR is an ill-posed problem since multiple possible HR images correspond to a single LR image.

In classical SR problems, the LR image is assumed to be the bicubically downsampled version of the HR image with known or small degradations. However, complicated image degradations are inevitable in real-world scenarios, and image blur is a common type of image degradation. During the image capturing process, the LR image may be degraded by various blur effects, such as motion blur, resulting from camera or scene motion during the exposure. Therefore, upscaling LR images with the SR technique will cause distorted HR results. So, it is essential to effectively combine deblurring and SR techniques to address the motion blurry image super-resolution problem. The motion deblurring problem is also highly ill-posed as multiple possible deblurred images correspond to a single motion blurry image. In this work, the problem of restoring the HR sharp image from a given motion blurry LR image is addressed.

The image degradation process consists of two parts for the motion blurry image super-resolution problem: blurring and downsampling. So, to tackle this joint problem, both deblurring and SR problems should be solved. During the last decade, deep convolutional neural networks and transformer-based [1] architectures have significantly advanced the image deblurring [2-7] and SR [8-12] techniques. Although both the image deblurring [2-7] and SR [8-12] methods generate state-of-the-art results, naively cascading them sequentially does not reconstruct blurry images well, as it is shown in [13]. There are several reasons: first, the error estimated from the first module will be magnified by the second module leading to error accumulation. Second, these two tasks are correlated, and it is sub-optimal to employ the feature extraction and image reconstruction phases twice. The features extracted from the deblurring module can be reused in the SR module to reconstruct the spatial details of the image.

Several recent methods jointly solve the motion blurry image super-resolution problem. Zhang et al. [14] proposed a dual-branch architecture to extract deblurring and SR features parallel and fuse them by the recursive gate module. However, the proposed deblurring and SR modules extract independent features, leading to sub-optimal results when blur is significant. Shoyan et al. [13] propose a single branch architecture by reusing the features extracted by the hierarchical layers of MPRNet [3] for image super-resolution. They refine the features extracted by MPRNet [3] and propagate them to the reconstruction module. However, their reconstruction module is not large enough to fully reuse the features extracted from the deblurring module.

Recently, the NTIRE 2021 Image Deblurring Challenge [15] was held where the Image Deblurring track 1 (low resolution) addresses the joint image deblurring and super-resolution problem. Several methods were proposed under the competition track 1 in the challenge. Bai et al. [16] developed a cascaded non-local residual network (CNLRN). The deblurring subnetwork is based on encoder-decoder architecture like SRN [2]. As a super-resolution subnetwork, they develop a non-local residual network to increase the resolution of the reconstructed features progressively. They propose a non-local block based on the self-attention mechanism [1] and use it in the SR subnetwork, thus achieving the 3rd PSNR [17] / SSIM [18] / LPIPS [19] scores in the NTIRE 2021 challenge [15]. However, as shown in Fig. 1, their method fails to remove the complex motion blur and recover enough sharpness for relatively small structures present in the LR blurry image. Also, the proposed network has a large number of parameters (~81M) and, therefore, has high computational complexity. Xi et al. [20] proposed a pixel-guided dual-branch attention network (PDAN). They design a residual spatial and channel attention module (RSCA) for feature extraction. They propose a Hard Pixel Example Mining (HPEM) loss to pay more attention to pixels with complicated degradation. Their method achieves the 2nd PSNR/SSIM scores in the NTIRE 2021 challenge [15]. However, the network performance mainly depends on the deep architecture and synthesized dataset. The network has about 61M parameters and considerable computational complexity as [16]. Also, the source code, pre-trained model, and network implementation details, such as the number of blocks used in feature extraction, reconstruction, and deblurring modules, are not publicly available, making their results non-reproducible. Xu et al. [21] proposed an enhanced deep pyramid network (EDPN). They adjust a

video restoration architecture to the blurry image super-resolution problem. Five replicated images are generated from the input image and fed into the network to extract the degraded LR image's self-scale and cross-scale similarity information. Their proposed pyramid progressive transfer (PPT) module performs feature alignment on the replicated images while the pyramid self-attention (PSA) module fuses the aligned features. The proposed network achieves the 1st PSNR/SSIM/LPIPS scores in the NTIRE 2021 challenge [15]. However, no pre-trained model is available for reproducing the results. Also, the publicly available source code is one of the ablation studies and has inconsistencies with some implementation details described in the paper.

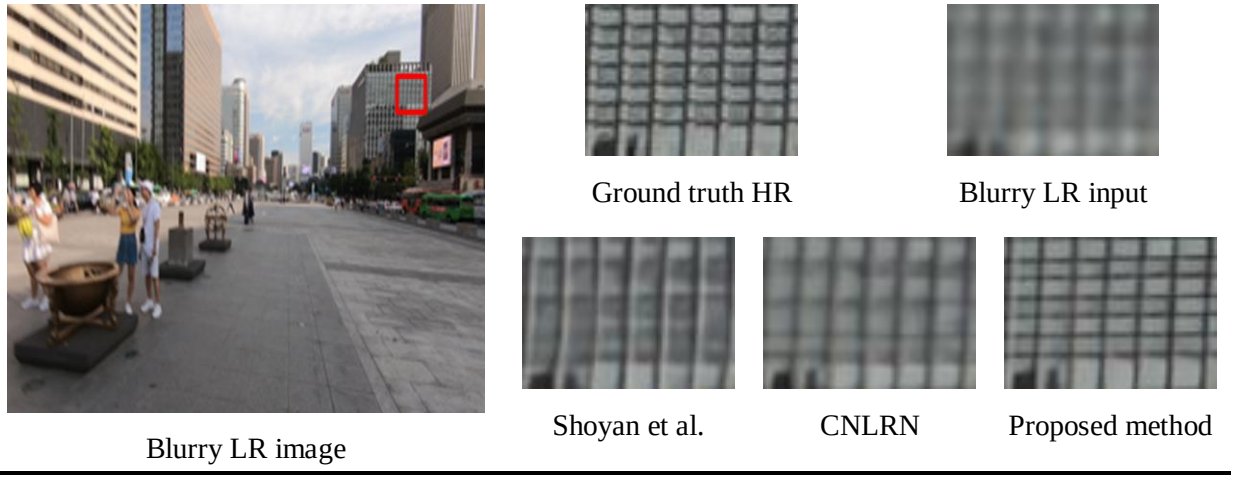

Fig. 1. Qualitative comparison between Shoyan et al. [13], CNLRN [16], and the proposed method. Please zoom in for the best view.

To tackle the problems mentioned above, a single branch architecture is proposed to effectively estimate the HR sharp image from the given motion blurry LR image. As shown in Fig. 1 (as well as in the Results section), in contrast to the other methods, the proposed method removes the complex motion blur from the small structures full of edges and generates an HR sharp image closer to the ground truth. Compared to the architectures suggested in [16] and [20], the proposed method has fewer parameters and, therefore, less computational complexity (see the Results section). Since MPRNet [3] generates contextually and spatially enriched features in its 3rd stage, the proposed architecture reuses these deblurring features in the reconstruction module to recover the HR sharp image. The reconstruction module is designed based on multi-scale channel attention modules (MS-CAM) [22] to process the small structures better. To train the whole network, a simple and effective training strategy is employed where 1) firstly, the deblurring module is trained, then 2) the SR module is trained using the pre-trained frozen deblurring module, and finally, 3) the deblurring module is unfrozen for joint end-to-end training.

The contributions of this paper are as follows:

- An end-to-end single branch architecture is proposed to effectively reconstruct the HR sharp image from the given motion blurry LR image.
- A reconstruction module is designed based on MS-CAM blocks [22] to better process the small structures in the LR blurry image.
- A simple and effective training strategy is exploited where the pre-trained frozen deblurring module is used as a feature extractor to train the SR module. Then the whole network is jointly trained with the unfrozen deblurring module.

The source code and the pre-trained model are publicly available at https://github.com/misakshoyan/joint-motion-deblur-and-sr.

2. Related Work

This section briefly reviews the motion deblurring, super-resolution, motion blurry super-resolution techniques and the related attention mechanism.

Motion deblurring. The motion deblurring problem for dynamic scenes is highly ill-posed since the blur kernel is spatially varying and unknown (non-uniform blind deblurring problem). Conventional blind motion deblurring methods jointly estimate the blur kernel and sharp image by solving computationally expensive optimization problems [23, 24]. Simplified priors are employed to regularize the solution space, such as dark channel prior [24], total-variation prior [23], etc. However, the simplified assumptions on the blur model make these techniques non-generalizable for real-world applications.

Recently, deep convolutional neural networks (CNN) have achieved significant success in image deblurring [2-7]. Since the blur kernel estimation is not practical in real-world applications, recent methods directly map the blurry image to the corresponding sharp image by designing an end-to-end image deblurring architecture. Nah et al. [6] proposed a deep multi-scale CNN called DeepDeblur to mimic the conventional coarse-to-fine optimization approaches. Tao et al. [2] proposed a scale recurrent network (SRN) and improved the DeepDeblur [6] network by sharing the network weights at different scales where each scale employs an encoder-decoder architecture. Zamir et al. [3] proposed a multi-stage progressive image restoration architecture called MPRNet. They employ an encoder-decoder subnetwork in the first and second stages to learn contextually enriched features due to large receptive fields. In contrast to the first and second stages, the last stage does not contain a downsampling operation. It employs a single-scale pipeline to generate spatially enriched features via channel attention blocks (CAB) [3]. A supervised attention module (SAM) is introduced to control the information flow between the stages, which refines the features before propagating them to the next stage. The MPRNet achieves competitive results in motion deblurring, image denoising, and deraining problems. Chu et al. [7] further improved the MPRNet results by exploring the statistics distribution inconsistency between training (with image patches) and testing (with full-size image) phases. They argue that the operations, which aggregate the global spatial statistics along the entire spatial dimension, may lead to a statistics distribution shift between training and testing phases since different size images are used in these phases. This statement mainly refers to the global average pooling operation used in CAB. To address this issue, they propose a test-time local statistics converter (TLSC) mechanism to calculate the mean value in a local window for each pixel rather than calculating a single mean value for the entire spatial dimension. They replace the global average pooling layer of CAB with TLSC-based local average pooling layers for each pixel at the test-time without re-training or fine-tuning the network. Zamir et al. [4] proposed a transformer-based [1] encoder-decoder architecture called Restormer. They calculate the self-attention [1] across channels rather than the spatial dimension, thus implicitly encoding the global context with linear complexity.

Super-Resolution. Like the motion deblurring problem, the SR problem is also highly ill-posed. Early approaches employ interpolation techniques such as bicubic and bilinear interpolations [25, pp. 77-78]. Some methods rely on other techniques such as neighbor embedding [26], sparse coding [27], etc. CNN-based methods [8-10] have achieved unprecedented success in single image SR during the last decade. Zhang et al. [8] designed a very deep residual channel attention network (RCAN). They focus on learning high-frequency information by employing residual in residual structure with long and short skip connections to bypass the abundant low-frequency information. The channel attention (CA) mechanism is proposed to model the interdependencies among feature channels inspired by the Squeeze and Excitation mechanism [28]. They design residual channel attention block (RCAB) using Conv-RELU-Conv structure followed by CA block and construct the network based on RCABs. Dai et

al. [9] designed a second-order attention network (SAN) by introducing second-order channel attention (SOCA) block to learn the channel-wise feature interdependencies better and to focus on more informative features using second-order channel statistics. They exploit share-source skip connections to bypass more abundant low-frequency information present in the LR image. Niu et al. [10] proposed a holistic attention network (HAN) based on RCAN [8]. The proposed layer attention module (LAM) allows the network to learn the interrelationships between features of different layers and focus on more informative layers. A channel-spatial attention module (CSAM) is introduced to learn the inter-channel and intra-channel dependencies for the last layer of the network.

However, the discussed SR methods assume known or small degradations and amplify the blur present in the LR image, as shown in [13]. Therefore, the SR network needs also to incorporate motion deblurring techniques.

Joint motion deblurring and super-resolution. Several recent methods solve the motion blurry image super-resolution problem by incorporating motion deblurring and SR techniques in a unified network. Zhang et al. [14] proposed a gated fusion network (GFN) to extract deblurring and SR features separately by designing a dual-branch architecture. The gate module fuses these features, and the reconstruction module generates an HR sharp image using the fused features. However, in their design, the SR branch operates on the blurry LR image, which limits the reconstruction ability of the network to generate sharp HR results, as shown in [13].

In contrast to GFN, Shoyan et al. [13] proposed a single branch architecture. They suggest reusing the features extracted by the hierarchical layers of MPRNet [3] since it generates contextually and spatially enriched features in its deep layers. These features are then refined and fed to the reconstruction module to generate the four times upscaled sharp HR image. However, it seems that their reconstruction module could be a bit larger to reuse the features extracted from the deblurring module fully.

Several methods were proposed in the scope of NTIRE 2021 Image Deblurring Challenge [15] under the Image Deblurring track 1 (low resolution) to solve the joint motion deblurring and SR problem. Bai et al. [16] cascaded the deblurring and super-resolution modules in a unified network. The deblurring module employs an encoder-decoder architecture like SRN [2], without a recurrent mechanism. As a super-resolution subnetwork, they develop a non-local residual network that contains RCABs [8] and self-attention-based [1] non-local blocks. The non-local block aims to model the global information for residual blur removal. A gradient loss function is developed to preserve the edges of the reconstructed HR image. They also propose a progressive upsampling mechanism to increase the resolution of the reconstructed features progressively and achieve the top-3 PSNR [17] / SSIM [18] / LPIPS [19] scores in the low-resolution track 1 of the NTIRE 2021 challenge [15]. However, as shown in Fig.1 (see also the Results section), this method fails to recover enough sharpness in its HR output when the relatively small structures of the LR blurry image are full of edges. Also, the non-local blocks and progressive upsampling mechanism have considerable computational complexity due to the self-attention mechanism [1]. In addition, the deblurring module has about 70M parameters because of the large convolutional kernel size.

Xi et al. [20] proposed a pixel-guided dual-branch attention network. The proposed residual spatial and channel attention module aims to better extract informative features by fusing cross-channel and spatial information. The deblurring module is based on an encoder-decoder architecture that employs residual blocks. Unlike the other architectures, the deblurring module is used only in the training stage for computational efficiency and helps the network extract and learn more useful deblurring information. The SR module is mainly composed of convolutional layers and pixel shuffling layers [29]. They argue that some pixels of the LR degraded image may contain a large amount of blur and downsampling degradation. In contrast, the other pixels may be only affected by downsampling, so they propose the Hard Pixel Example Mining loss to

pay more attention to pixels with complicated degradation. Their network achieves top-2 PSNR/SSIM scores on track 1 of the NTIRE 2021 challenge [15]. However, the network has considerable computational complexity as [16] since it has about 61M parameters. Unlike other methods, the network's performance depends on the additionally synthesized dataset (~72K images) used to fine-tune the network. In addition, the source code and the implementation details of the network (such as the number of blocks used in feature extraction, reconstruction, and deblurring modules) are not publicly available for researchers to reproduce the results.

Xu et al. [21] adjusted a video restoration architecture to the blurry image super-resolution problem by feeding five replicated images into the network to fully exploit the degraded LR image's self-scale and cross-scale similarity information. Their proposed pyramid progressive transfer (PPT) module employs a pyramid structure and aims to generate attention masks to progressively transfer the self-similarity information. The pyramid self-attention (PSA) module aggregates and reweights the transferred features with a pyramid structure. Their proposed network achieves the best PSNR/SSIM/LPIPS scores in the NTIRE 2021 Image Deblurring Challenge [15]. However, the pre-trained model is not provided for researchers. The publicly available source code is one of the ablation studies of EDPN that has some inconsistencies with the paper. Therefore, the method results are not reproducible, like [20].

Attention mechanism. The attention mechanism aims to mimic the human visual perception to pay attention to the informative part of the input. The self-attention [1] mechanism allows modeling the global dependencies between each word in a sentence or pixel in the image. Hu et al. [28] proposed the Squeeze and Excitation block (SE) to model the interdependencies between the feature channels. Dai et al. [9] designed the second-order channel attention block by replacing the global average pooling operation of the SE block with the global covariance pooling operation [9] to capture higher-order feature statistics for more discriminative representations. To solve the problems of scale variation and small objects, Dai et al. [22] designed the multi-scale channel attention module (MS-CAM) for aggregating both the local and global feature contexts within the channel attention. The global average pooling operation employed in the SE block emphasizes the globally distributed large objects while potentially ignoring the locally distributed small objects. Therefore, in addition to the global branch, a local branch is added into SE to simultaneously aggregate the global and local contexts within a multi-scale channel attention mechanism.

3. Proposed Method

This work proposes a single-branch network architecture to solve the joint motion deblurring and SR problem. Unlike the existing methods, the proposed method successfully reconstructs the relatively small degraded structures (see the Results section). The network reuses the contextually and spatially enriched features generated by MPRNet [3] for image super-resolution. The SR module is designed by employing residual in residual structure based on multi-scale channel attention modules (MS-CAM [22]), which simultaneously aggregate the local and global feature contexts within the channel attention to better process the small objects.

Network Architecture. To solve the motion deblurring and SR problems jointly, the network should extract both contextually and spatially informative features. The encoder-decoder architectures [2, 4] are effective in encoding the contextually informative features, which are helpful for the motion deblurring problem. The single-scale pipelines [8-10] extract spatially-enriched features that are informative for the SR task. The MPRNet [3] naturally offers a twofold functionality since it employs both encoder-decoder and single-scale pipelines.

The proposed network consists of three main parts: deblurring, feature transformation, and super-resolution (see Fig. 2).

Deblurring Module. The deblurring module is based on MPRNet [3] and employs multi-stage progressive architecture with three stages (see Fig. 2). An encoder-decoder architecture is employed in the first and second stages of MPRNet based on U-Net architecture [30] to generate contextually informative features due to the large receptive field. Each scale of the encoder and decoder networks uses 2 channel attention blocks (CAB) [3]. Also, each skip connection between encoder and decoder uses 2 CABs. The CAB block incorporates Conv-RELU-Conv structure and Squeeze-and-Excitation block (SE) [28] with the residual connection. It reweights the input feature map by emphasizing more informative channels. The 3rd stage employs a single-scale original-resolution subnetwork (ORSNet) [3] and aims to generate spatially informative features without downsampling and upsampling operations. It contains 3 original-resolution blocks (ORB) [3]. Each ORB is a residual group with 8 CABs followed by a convolution layer.

The supervised attention module (SAM) [3] and the cross-stage feature fusion (CSFF) [3] mechanism are employed to control the information flow between two consecutive stages. The SAM module uses the ground-truth supervisory image. It generates attention maps to suppress the less informative features of the current stage and to propagate only the more informative ones to the next stage. The CSFF mechanism propagates the intermediate contextualized features of the previous stage to the next stage. It aims to compensate for the information loss due to repeated upsampling and downsampling operations performed in the first and second stages. Thus, the SAM module and CSFF mechanism allow the ORSNet to combine the generated spatially informative features with the contextually informative features of the previous stage and allow the MPRNet to produce both contextually and spatially enriched features in its 3rd stage.

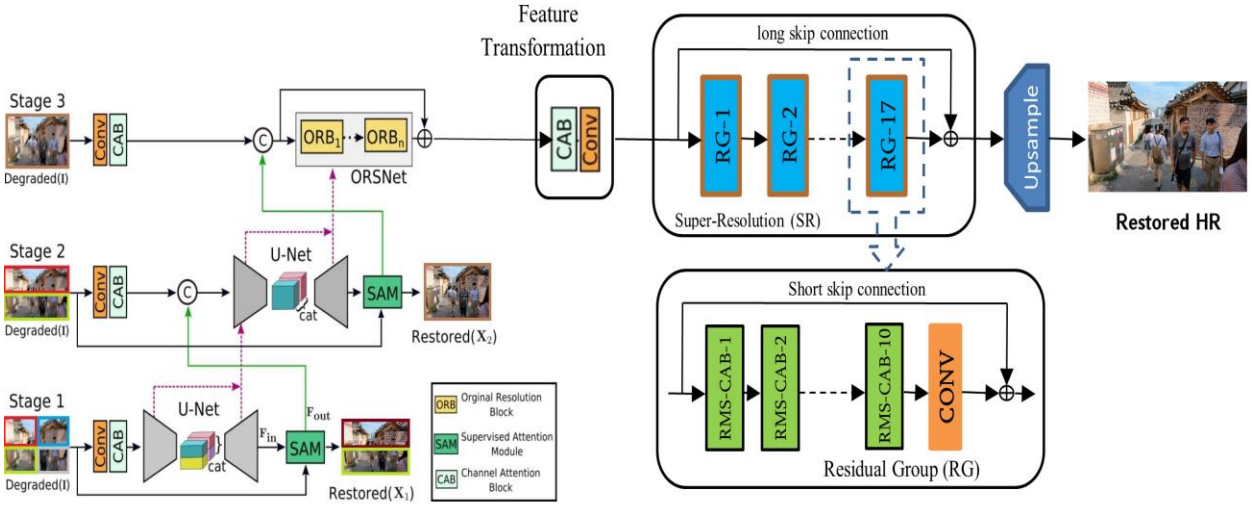

Fig. 2. The architecture of the proposed method. Please, zoom in for the best view.

Feature Transformation Module. The feature transformation module refines the contextually and spatially enriched features generated by MPRNet [3] before propagating them to the super-resolution module. It contains a CAB block followed by a convolutional layer. A CAB block is applied on MPRNet generated features to emphasize the more informative features for SR and suppress the less informative ones. Since the third stage of MPRNet operates on 128 channels, a convolutional layer transforms the CAB-generated features by decreasing the number of channels from 128 to 64. It aims to achieve a trade-off between computational complexity and accuracy for the SR.

Super-resolution Module. The super-resolution module takes the transformed features as input and employs a single-scale pipeline to reconstruct the HR sharp image. The image SR problem can be treated as a process to recover the high-frequency information (like regions full of edges) as much as possible since the low-frequency information is less informative (like uniform regions). Thus, the abundant low-frequency information can be forwarded to the final HR output with relatively less processing. Inspired by RCAN [8], a residual-in-residual (RIR) structure is exploited to bypass the abundant low-frequency information via long and short skip connections and make the network focus on learning the high-frequency information. The RIR structure stacks 17 residual groups (RG), where each residual group contains 10 residual multi-scale channel attention blocks (RMS-CAB) followed by a convolution layer (see Fig. 3). After the RIR structure, the pixel-shuffling layer [29] follows to upscale the spatial resolution of the features four times. Then a convolutional layer is used to generate a colored image from 64 channels.

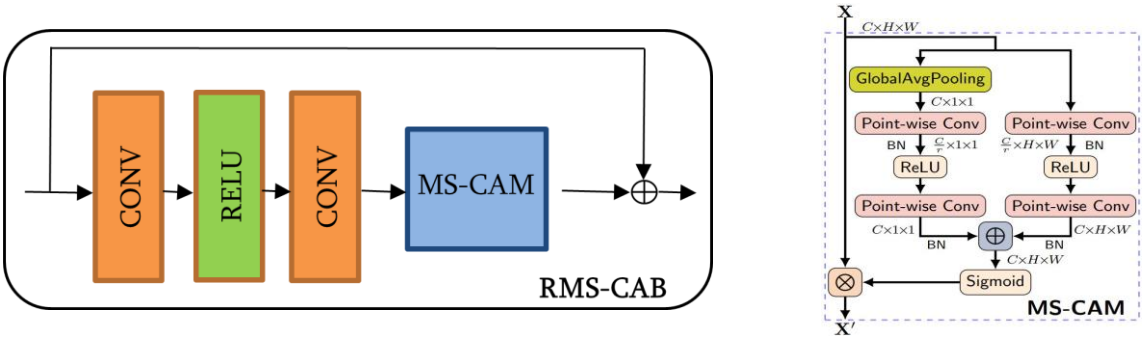

Fig. 3. The RMS-CAB block and MS-CAM [22] block. Please, zoom in for the best view.

The employed RMS-CAB block is inspired by CAB [3]. It consists of a Conv-RELU-Conv structure and MS-CAM [22] block with the residual connection. Different from CAB, it employs MS-CAM [22] block instead of SE block [28] to aggregate both the local and global feature contexts. The SE and MS-CAM blocks aim to capture the interdependencies among feature channels and reweight them by amplifying the more informative channels and suppressing the less informative ones. In contrast to the SE block, the MS-CAM block captures the information from the globally distributed large and locally distributed small objects, while the SE block ignores the signals present in small objects. The MS-CAM block employs global and local branches (see Fig. 3) to implement the channel attention in two scales. The global branch applies a global average pooling operation and calculates the mean value for each channel to capture the channel-wise feature statistics globally. Then two point-wise convolutions follow with RELU activation between them. Unlike the global branch, the local branch operates on the original spatial resolution and does not exploit the global average pooling operation. The features generated by two branches are aggregated via broadcasting addition. Then a sigmoid activation follows, and the final features are element-wise multiplicated with the input.

Training Strategy and Loss Function. The whole network can be trained jointly from scratch, but this approach causes sub-optimal results. The reason is that for the most part of the joint training, the deblurring module will generate blurry results, which will hamper the training process of the SR module. As a result, the network will not learn the exact mapping between blurry LR and sharp HR images. Since it would be beneficial for the SR module to operate on deblurred features, a simple and effective training strategy is exploited, consisting of three phases.

In the first training phase, only the deblurring module (MPRNet) is trained. The MPRNet generates LR deblurred image at each stage, denoted as ($\hat{L}_1$, $\hat{L}_2$, $\hat{L}_3$). Since, in the proposed

architecture, the third stage of the MPRNet does not produce a deblurred image, a convolutional layer is added to generate the $\hat{L}_3$ image for the training phase only. The deblurring loss is defined as

$$\mathcal{L}_{\mathrm{DB}} = \sum_{s=1}^{3} \mathcal{L}_{\mathrm{char}}\left(\hat{\mathrm{L}}_{\mathrm{s}}, \mathrm{L}\right) + \lambda_{edge} * \mathcal{L}_{\mathrm{edge}}(\hat{\mathrm{L}}_{\mathrm{s}}, \mathrm{L}),$$

$$\mathcal{L}_{\mathrm{char}} = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \sqrt{||\hat{\mathrm{L}}_s^i - \mathrm{L}^i||^2 + \varepsilon^2},$$

$$\mathcal{L}_{\mathrm{edge}} = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \sqrt{||\Delta\hat{\mathrm{L}}_s^i - \Delta\mathrm{L}^i||^2 + \varepsilon^2},$$

where $\mathcal{L}_{\mathrm{DB}}$ denotes the deblurring loss, N is the number of training images, and L is the bicubic downsampled version of the ground-truth HR image. As in [3], $\mathcal{L}_{\mathrm{char}}$ and $\mathcal{L}_{\mathrm{edge}}$ denote the Charbonnier loss [31] and the Edge loss [3], Δ denotes the Laplacian operator. ε and λ_{edge} were empirically set to 10^{-3} and 0.05, respectively.

In the second training phase, only the SR module is trained (including the transformation module) by employing the pre-trained frozen deblurring module. It is beneficial for the SR module since, in contrast to joint training from scratch, it starts to operate on fully processed deblurred features and does not have to deal with blurry features now. To preserve the structural details, like edges, the combination of pixel-wise and gradient losses ($\mathcal{L}_{\mathrm{pixel}}$ and $\mathcal{L}_{\mathrm{grad}}$) is employed as in [16]:

$$\mathcal{L}_{\mathrm{pixel}} = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left\| \hat{\mathrm{H}}^i - \mathrm{H}^i \right\|_1,$$

$$\mathcal{L}_{\mathrm{grad}} = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left\| \nabla\hat{\mathrm{H}}^i - \nabla\mathrm{H}^i \right\|_1,$$

$$\mathcal{L}_{\mathrm{SR}} = \mathcal{L}_{\mathrm{pixel}} + \lambda_{grad}\mathcal{L}_{\mathrm{grad}},$$

where $\hat{\mathrm{H}}$ and H are the reconstructed and ground-truth high-resolution images, respectively. $\mathcal{L}_{\mathrm{SR}}$ denotes the SR loss, ∇ is the image gradient operator and λ_{grad} is set to 0.1 as in [16].

In the last training phase, the deblurring module is unfrozen, and the whole network is trained jointly. The deblurring loss is employed only in the first and second stages of the MPRNet. This trick aims to force the third stage of MPRNet to pay more attention to generating the SR features during the last phase of the training. The following combination of loss functions is exploited

$$\mathcal{L} = \mathcal{L}_{\mathrm{SR}} + \alpha\mathcal{L}_{\mathrm{DB}},$$

where α was empirically set to 0.5 as in [13].

Implementation Details. The REDS dataset [32] is used to train the proposed network. It contains 24.000 training images, 3000 validation images, and 3000 testing images. The validation dataset is used for quantitative evaluations since the ground-truth images of the testing set are not available. The size of the LR blurry and ground-truth HR images is equal to 320x180 and 1280x720, respectively.

The proposed method is trained with the three-phase training strategy as discussed above. The training patch size for the blurry LR and ground-truth HR images is 64x64 and 256x256, respectively, which are randomly cropped from the training set. Randomly horizontal and vertical flips are applied combined with rotation for data augmentation. The Adam optimizer [33] is used with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The initial learning rate and the batch size for the three phases of training are set to $(10^{-4}, 10)$, $(10^{-4}, 3)$, and $(10^{-5}, 3)$, respectively. During the training of each phase, the learning rate was gradually modified to obtain the best results. The network is trained and evaluated on a single 'NVIDIA GeForce GTX 1660 Ti with Max-Q Design' GPU.

At the test time, the TLSC mechanism [7] is applied to the whole network to address the issue of statistics distribution shift between the training and testing phases. The local window size of TLSC was empirically set to 96x96. The proposed method takes about 1.1 seconds to recover a 1280x720 size HR image from the 320x180 size LR blurry input.

4. Results

The proposed network is evaluated on the REDS validation dataset [32] both quantitatively and qualitatively. The PSNR [17] / SSIM [18] scores are calculated using the computation codes released by [16] for a fair comparison.

Table 1 summarizes the quantitative comparison results between the existing methods on the Val300 [16] dataset derived from the REDS validation set by sampling each tenth image from the validation set.

Table 1: Quantitative comparison results on the Val300 dataset [16]. * denotes the results cited from [16], # denotes the self-ensemble strategy [12].

Methods	Bicubic*	GFN*	RCAN*	[MPRNet + RCAN]*	[SRN + RCAN]*	Shoyan et al	CNLRN	*Proposed method*
PSNR	23.848	26.635	27.338	27.550	27.610	27.164	27.770 / 27.922#	*27.620 / 27.740#*
SSIM	0.6481	0.7447	0.7661	0.7740	0.7745	0.7610	0.7784 / 0.7813#	*0.7735 / 0.7760#*

As shown in Table 1, the PSNR score of the SR method RCAN [8] is at least 0.29dB less than the proposed method since the blur degradation hampers the performance of SR techniques. The existing SR methods [8-10] employ only a single-scale pipeline, which fails to extract the contextual information to remove the blur degradation. The joint method GFN [14] does not perform well since the SR branch operates on the blurry LR image. The joint method proposed by Shoyan et al. [13] generates promising results (see Fig. 1, Fig. 4), but their reconstruction module is not large enough to fully reuse the contextually and spatially enriched features extracted from MPRNet [3]. The cascaded approaches of deblurring and SR methods (MPRNet [3] + RCAN [8], SRN [2] + RCAN [8]) still obtain less PSNR/SSIM scores than the proposed method since they employ the Squeeze-and-Excitation mechanism [28]. In contrast, the proposed method exploits the benefits of MS-CAM [22] blocks to better process the relatively small structures present in the LR blurry image. The joint method CNLRN [16] obtains higher PSNR/SSIM results than the proposed method. However, the CNLRN [16] fails to reconstruct enough sharpness for relatively small structures in the LR blurry image and generates smoothed

texture details, as shown in Fig. 4. The self-ensemble strategy [12] is also employed to increase the performance of the proposed method by running the model on 8 augmented LR images followed by inverse transformation and averaging, resulting in 27.740/0.7760 dB PSNR/SSIM scores, respectively.

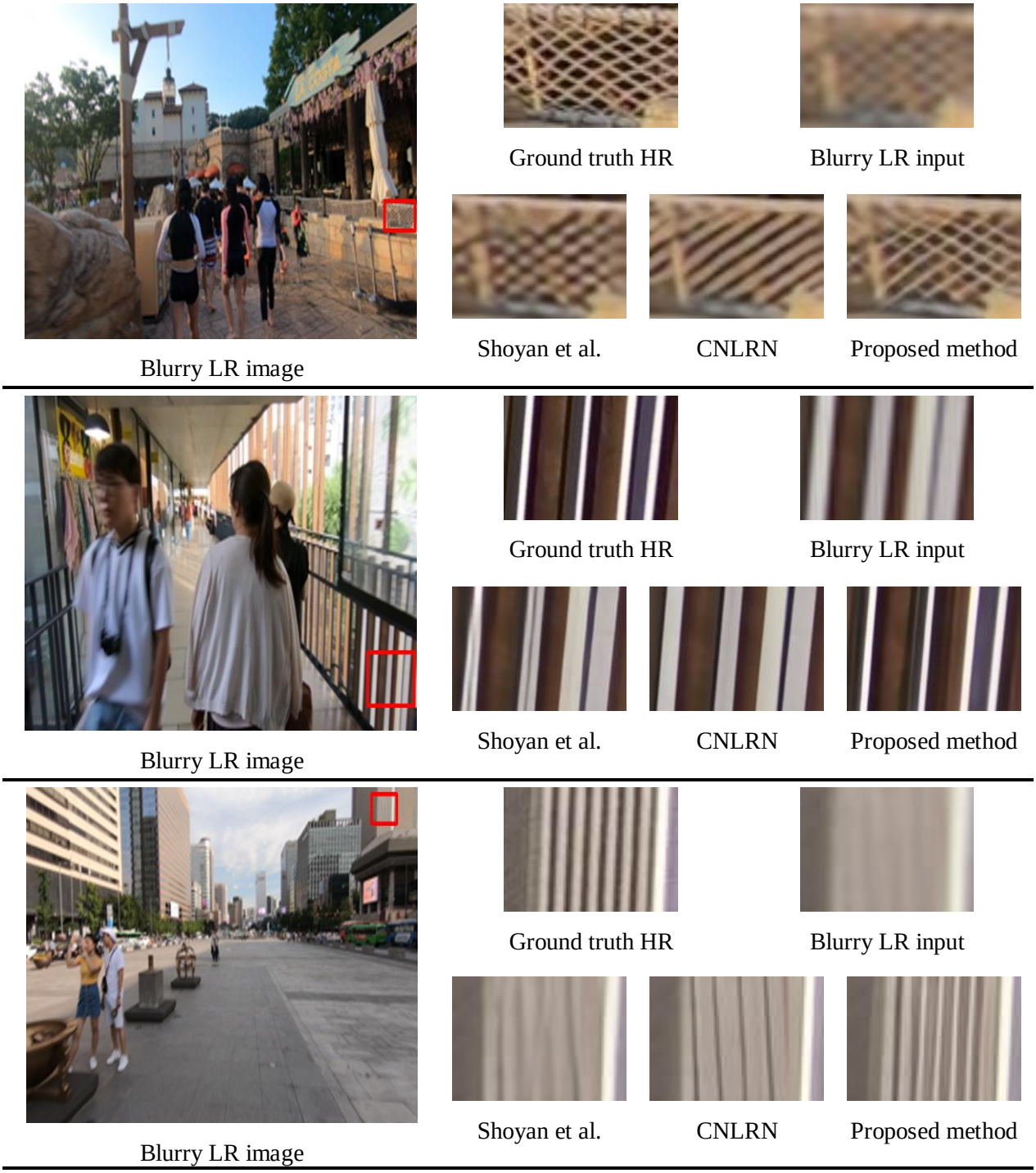

Fig. 4. Qualitative comparison results on the validation dataset. Please, zoom in for the best view.

Table 2: Quantitative comparison results on the REDS [32] validation dataset.

Methods	CNLRN	PDAN	EDPN	*Proposed method*
PSNR	27.828	27.890	28.010	*27.660*
SSIM	0.7794	0.7798	0.8203	*0.7745*
#Params	83.06M	61.00M	13.34M	*34.12M*

Table 2 summarizes the quantitative comparison results for the proposed method and three top-ranked methods in NTIRE 2021 [15] challenge (CNLRN [16], PDAN [20], EDPN [21]). The methods were evaluated on the REDS [32] validation set. The results of PDAN [20] and EDPN [21] were cited from their papers. As shown in Table 2, the proposed method obtains less PSNR/SSIM scores than the other methods, but it has fewer parameters (~34M) and, therefore, less computational complexity compared to CNLRN (~83M) and PDAN (~61M). It makes the model more applicable in scenarios when computational resources are limited. It should be mentioned that the results of PDAN are not reproducible since the source code, pre-trained model, and network implementation details are not publicly available. It also refers to EDPN since no pre-trained model is provided, while the publicly available source code is one of the authors' ablation studies.

The qualitative comparison results on the REDS validation set are demonstrated in Fig. 4. As it is shown, the CNLRN [16] and the method proposed by Shoyan et al. [13] fail to recover the small structures present in the LR blurry image. Instead, the proposed method successfully removes the complex motion blur from the relatively small structures and reconstructs fine texture details. As shown, unlike the other methods, the HR image generated by this method is closer to the ground-truth HR image.

5. Ablation Study

The ablation experiments were performed on the Val300 dataset. Table 3 summarizes the effect of different modifications in terms of PSNR [17] / SSIM [18] metrics on the RGB channel for the proposed method. The baseline model employs only the deblurring and transformation modules followed by upsampling layer [29]. As a Model-1, the SR module is added to the baseline model with 12 RGs, each containing 10 RMS-CAB blocks followed by a convolution layer.

Table 3: Ablation study on the Val300 dataset [16].

Modifications	Models			
	Baseline	**Model-1**	**Model-2**	**Model-3**
Baseline	✓	✓	✓	✓
120 RMS-CAB		✓		
170 RMS-CAB			✓	✓
TLSC				✓
PSNR	27.098	27.453	27.501	**27.620**
SSIM	0.7530	0.7696	0.7713	**0.7735**

This change brings ~0.36dB improvement in terms of PSNR. When the number of RGs is increased to 17, the PSNR score is improved from 27.453dB to 27.501dB (Model-2). Finally, the TLSC mechanism [7] is employed on the whole network (Model-3) at the test time, achieving a 27.620dB score in terms of PSNR.

The models were trained with the proposed three-phase training strategy by gradually modifying the learning rate to obtain the best results.

To show the effect of the proposed training strategy, the Model-1 is trained jointly from scratch. It achieves 27.371dB in terms of PSNR, about 0.08dB less than the same model trained with the proposed three-phase training strategy.

6. Conclusion

This paper proposes an end-to-end single branch architecture to reconstruct a sharp HR image from the given motion blurry LR image. The proposed method reuses the contextually and spatially enriched features extracted from the MPRNet [3] in the super-resolution subnetwork. The super-resolution subnetwork is designed based on MS-CAM [22] blocks. A three-phase training strategy is exploited where a pre-trained frozen deblurring module is used as a feature extractor to train the SR module. The deblurring module is unfrozen in the last phase of the training. Experiments show that, unlike the other methods, the proposed method successfully removes the complex motion blur from the relatively small structures and reconstructs fine texture details. The source code and the pre-trained model are available at https://github.com/misakshoyan/joint-motion-deblur-and-sr.

References

[1] A. Vaswani et al. “Attention Is All You Need”, *arXiv preprint arXiv:1706.03762*, 2017.

[2] X. Tao et al., “Scale-recurrent network for deep image deblurring”, *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, pp. 8174–8182, 2018.

[3] S. W. Zamir et al., “Multi-Stage Progressive Image Restoration”, *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, USA, pp. 14816-14826, 2021.

[4] S. W. Zamir et al. “Restormer: Efficient Transformer for High-Resolution Image Restoration”, *arXiv preprint arXiv:2111.09881*, 2021.

[5] X. Mao et al. “Deep Residual Fourier Transformation for Single Image Deblurring”, *arXiv preprint arXiv:2111.11745*, 2021.

[6] S. Nah, T. H. Kim, and K. M. Lee, “Deep multi-scale convolutional neural network for dynamic scene deblurring”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, pp. 257-265, 2017.

[7] X. Chu et al. “Revisiting Global Statistics Aggregation for Improving Image Restoration”, *arXiv preprint arXiv:2112.04491*, 2021.

[8] Y. Zhang et al., “Image super-resolution using very deep residual channel attention networks”, *Proceedings of European Conference on Computer Vision (ECCV)*, Munich, Germany, pp. 294-310, 2018.

[9] T. Dai et al., “Second-Order Attention Network for Single Image Super-Resolution”, *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, USA, pp. 11057-11066, 2019.

[10] B. Niu et al., "Single image super-resolution via a holistic attention network", *Proceedings of European Conference on Computer Vision (ECCV)*, Glasgow, UK, pp. 191–207, 2020.
[11] J. Liang et al., "SwinIR: Image Restoration Using Swin Transformer", *Proceedings of IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, Canada, pp. 1833-1844, 2021.
[12] B. Lim et al., "Enhanced deep residual networks for single image super-resolution", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, USA, pp. 1132-1140, 2017.
[13] M. Shoyan et al., "Single Image Joint Motion Deblurring and Super-Resolution", *Proceedings of 13th International Conference on Computer Science and Information Technologies (CSIT)*, Yerevan, Armenia, pp. 182-186, 2021.
[14] X. Zhang et al. "Gated fusion network for degraded image super resolution", *International Journal of Computer Vision*, vol. 128, no. 6, pp. 1699-1721, 2020.
[15] S. Nah et al., "NTIRE 2021 Challenge on Image Deblurring", *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Nashville, USA, pp. 149-165, 2021.
[16] H. Bai et al, "Learning A Cascaded Non-Local Residual Network for Super-resolving Blurry Images", *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Nashville, USA, pp. 223-232, 2021.
[17] Wikipedia, (2014) The peak signal-to-noise ratio, [Online]. Available: https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio
[18] Z. Wang et al., "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.
[19] R. Zhang et al., "The unreasonable effectiveness of deep features as a perceptual metric", *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, pp. 586-595, 2018.
[20] S. Xi, J. Wei and W. Zhang, "Pixel-Guided Dual-Branch Attention Network for Joint Image Deblurring and Super-Resolution", *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Nashville, USA, pp. 532-540, 2021.
[21] R. Xu et al., "EDPN: Enhanced Deep Pyramid Network for Blurry Image Restoration", *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Nashville, USA, pp. 414-423, 2021.
[22] Y. Dai et al., "Attentional Feature Fusion", *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, USA, pp. 3559-3568, 2021.
[23] T. H. Kim, B. Ahn and K. M. Lee, "Dynamic Scene Deblurring", *Proceedings of IEEE International Conference on Computer Vision*, Sydney, Australia, pp. 3160-3167, 2013.
[24] J. Pan et al., "Blind image deblurring using dark channel prior", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 1628-1636, 2016.
[25] R. Gonzalez and R. Woods, Digital Image Processing, 4th ed., Pearson, New York, 2018.
[26] M. Bevilacqua et al. "Low-complexity single-image super-resolution based on nonnegative neighbor embedding", *Proceedings of British Machine Vision Conference (BMVC)*, Guildford, UK, paper 135, pp. 1-10, 2012.
[27] R. Timofte, V. D. Smet, and L. V. Gool. "A+: Adjusted anchored neighborhood regression for fast super-resolution", *Proceedings of Asian Conference on Computer Vision (ACCV)*. Singapore, Singapore, pp. 111-126, 2014.

[28] J. Hu, L. Shen and G. Sun, “Squeeze-and-Excitation Networks”, *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132-7141, 2018.
[29] W. Shi et al., “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 1874-1883, 2016.
[30] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation”, *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, pp. 234-241, 2015.
[31] P. Charbonnier et al, “Two deterministic half-quadratic regularization algorithms for computed imaging”, *Proceedings of 1st International Conference on Image Processing*, Austin, USA, pp. 168-172, 1994.
[32] S. Nah et al., “NTIRE 2019 challenges on video deblurring and super-resolution: Dataset and study”, *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, USA, pp. 1996–2005, 2019.
[33] D. Kingma and J. Ba, “Adam: A method for stochastic optimization”, *arXiv preprint arXiv:1412.6980*, 2014.

Պատկերից շարժման հետևանքով առաջացած շաղվածության հեռացում և պատկերի կետայնության բարձրացում՝ օգտագործելով շերտային ուշադրության բազմամասշտաբ մոդուլները

Միսակ Տ. Սհոյան

Հայաստանի Ազգային Պոլիտեխնիկական Համալսարան

e-mail: misakshoyan@gmail.com

Ամփոփում

Վերջին տասնամյակում խորը փաթույթային նեյրոնային ցանցերը զգալիորեն զարգացրել են պատկերի կետայնության բարձրացման մեթոդները՝ վերակառուցելով պատկերի կառուցվածքային և տարածական իրատեսական մանրամասներ։ Պատկերի կետայնության բարձրացման դասական խնդիրներում ենթադրվում է, որ ցածր կետայնության պատկերն ունի կետայնության նվազման որոշակի դեգրադացիա։ Սակայն, իրական սցենարներում, պատկերի բարդ դեգրադացիաներն անխուսափելի են և շարժման հետևանքով առաջացած շաղվածությունը պատկերի դեգրադացիայի տարածված տեսակ է, որն առաջանում է պատկերի նկարահանման գործընթացում՝ տեսախցիկի կամ տեսարանի շարժման հետևանքով։ Այս աշխատանքում առաջարկվում է լրիվ փաթույթային նեյրոնային

ցանց՝ տրված ցածր կետայնությամբ և շարժման հետևանքով առաջացած շաղվածություն պարունակող պատկերներից բարձր կետայնության հստակ պատկերներ վերակառուցելու համար: Շաղվածության հեռացման ենթացանցը հիմնված է բազմափուլային պրոգրեսիվ ճարտարապետության վրա, մինչդեռ կետայնության բարձրացման ենթացանցը նախագծելու համար օգտագործվում են շերտային ուշադրության բազմամասշտաբ մոդուլները: Կիրառվում է ուսուցանման պարզ և արդյունավետ եղանակ, որտեղ նախապես ուսուցանված և սառեցված շաղվածության հեռացման մոդուլն օգտագործվում է կետայնության բարձրացման մոդուլն ուսուցանելու համար: Ուսուցման վերջին փուլում շաղվածության հեռացման մոդուլը նույնպես ուսուցանվում է: Կատարված փորձերը ցույց են տալիս, որ ի տարբերություն մյուս մեթոդների, առաջարկվող մեթոդը վերակառուցում է համեմատաբար փոքր դետալները և կառուցվածքային մանրամասները՝ միաժամանակ հաջողությամբ հեռացնելով շարժման հետևանքով առաջացած բարդ շաղվածությունը: Իրականացման կոդը և նախապես ուսուցանված մոդելը հասանելի են https://github.com/misakshoyan/joint-motion-deblur-and-sr կայքում:

Բանալի բառեր՝ Շարժման հետևանքով առաջացած շաղվածության հեռացում, կետայնության բարձրացում, շերտային ուշադրություն:

Удаление размытости вызванной движением и увеличение разрешения одного изображения с использованием многомасштабных модулей внимания канала

Мисак Т. Сгоян

Национальный Политехнический Университет Армении

e-mail: misakshoyan@gmail.com

Аннотация

За последнее десятилетие глубокие сверточные нейронные сети значительно продвинули методы увеличения разрешения одного изображения, реконструируя реалистичные текстурные и пространственные детали. В классических задачах увеличении разрешения изображения предполагается, что изображение с низким разрешением имеет определенную деградацию понижения разрешении. Однако, в реальных сценариях, сложные деградации изображения неизбежны, и размытие изображения вызванное движением является распространенным типом деградации, которое происходит в процессе съемки изображения из-за движения камеры или сцены. В этой работе предлагается полностью сверточная нейронная сеть для реконструкции четких изображений с высоким разрешением из заданных размытых изображений вызванных движением с низким разрешением. Подсеть удаления размытости основана на

многоступенчатой прогрессивной архитектуре, в то время как подсеть увеличения разрешения разработана с использованием многомасштабных модулей внимания каналов. Используется простая и эффективная стратегия обучения, в которой предварительно обученный и замороженный модуль удаления размытости используется для обучения модуля увеличения разрешения. Модуль удаления размытости размораживается на последнем этапе обучения. Эксперименты показывают, что, в отличие от других методов, предлагаемый метод реконструирует относительно небольшие структуры и текстурные детали, успешно удаляя сложное размытие вызванное движением. Код реализации и предварительно обученная модель общедоступны по адресу https://github.com/misakshoyan/joint-motion-deblur-and-sr.

Ключевые слова: Удаление размытости вызванной движением, увеличение разрешения, внимание канала.

Mathematical Problems of Computer Science 56, 35–47, 2021.
doi: 10.51408/1963-0077

UDC 004.934

Deep Learning Approaches for Voice Emotion Recognition Using Sentiment-Arousal Space

Narek T. Tumanyan

Weizmann Institute of Science, Israel
e-mail: narek.tumanyan@weizmann.ac.il

Abstract

In this paper, we present deep learning-based approaches for the task of emotion recognition in voice recordings. A key component of the methods is the representation of emotion categories in a sentiment-arousal space and the usage of this space representation in the supervision signal. Our methods use wavelet and cepstral features as efficient data representations of audio signals. Convolutional Neural Network (CNN) and Long Short Term Memory Network (LSTM) architectures were used in recognition tasks, depending on whether the audio representation was treated as a spatial signal or as a temporal signal. Various recognition approaches were used, and the results were analyzed.
Keywords: Voice emotion recognition, Sentiment-arousal space, Spectral features, Speech sentiment classification.
Article Info: Received 19 July 2021; accepted 26 October 2021.

1. Introduction

In this work, we address the problem of emotion recognition from voice recordings. Recognizing emotion from voice can have various real-world applications, such as in recommendation systems, security systems, customer services, etc. Defining the recognition task formally, we want to come up with a model F, such that given a voice recording X in some representation, the model will give us a mapping $F(X) = y$, where y is some descriptor of the recognized emotion from the audio signal. Now, the question is, what space does y belong to? Is it discrete or continuous, and how are emotion values organized in this space? To answer these questions, we utilize a sentiment-arousal space described in the paper, which allows us to tackle the recognition task in different approaches, depending on how we use this space for defining the set of y values.

Previous methods for the voice emotion recognition problem include SVM-based classification algorithms [1], which also consider visual data of the facial expression of the speaker as an additional signal, as well as Deep Neural Network Extreme Learning method with an efficient performance on small datasets [2].

We use Mel Frequency Cepstral Coefficients (MFCC) and Continuous Wavelet Transforms (CWT) for representing audio signals in spectral features. Convolutional Neural Networks (CNN) and Long Short Term Memory Networks (LSTM) were used as deep learning model architectures.

2. Datasets

In our work, we used 3 databases of labeled voice recordings: Surrey Audio-Visual Expressed Emotion (SAVEE) [3], Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [4], and Toronto Emotional Speech Set (TESS) [5]. The databases are comprised of voice recordings of individuals who pronounce a statement with an exerted emotion, which is the label of the given voice recording. The emotion labels in the RAVDESS database are: "neutral", "calm", "happy", "sad", "angry", "fearful", "disgust", "surprised". TESS and SAVEE datasets have the same emotion labels except the "calm" one. The distribution of samples and labels of the databases are summarized in Table 2 and in Table 1.

Table 1: Number of voice recordings per emotion label across all databases.

Neutral	Calm	Sad	Fear	Anger	Surprises	Happiness	Disgust
616	192	652	652	652	652	652	652

Table 2: Summary of datasets used.

Database	Num of Recordings	Num of Actors	Emotion Labels
RAVDESS	1440	24	8
SAVEE	480	4	7
TESS	2880	2	7

3. Method

3.1 Feature Extraction

The first step in data preparation is resampling the voice recording signal in a certain sampling rate. As the signal in interest is human voice, which is known to lie in frequency ranges 4-10 Khz, we chose 22.05 Khz sampling rate. The resampled signal includes the human voice signal along with some possible frequency variations, which can be caused by possible pronunciation of high frequency sounds, such as fricatives. As a result, we obtain a temporal signal representation of the voice recording, which at a given time point shows the amplitude of air pressure oscillations from 0 frequency.

3.1.1 Fourier Representation

A temporal signal $x(t)$ can be represented as a combination of periodic functions of varying frequencies [6]

$$x(t) = \int_{-\infty}^{\infty} X(w) e^{jwt} dw,$$

where w denotes the frequency of the periodic function. Having the coefficients $X(w)$ is equivalent to having the original signal $x(t)$, and these coefficients are used as a representation of the signal in frequency domain. Such a representation is obtained by the Fourier Transform operation [6]. Discrete Fourier Transform (DFT) is the discrete equivalent of Fourier Transform, which we leverage for representing our discrete resampled signal $x[n]$ of length k in frequency space through coefficients / intensities $X[k]$ for each frequency k [7]:

$$X[k] = \sum_{n=1}^{K} x[n] e^{-i2\pi kn/N};\ 1 \leq k \leq K.$$

Usually, representing the entire discrete signal $x(t)$ with Fourier coefficients can result in loss of temporal resolution, since having a Fourier representation for the entire signal does not include changes of the signal in small temporal windows. For obtaining higher resolution in temporal domain, Short-Time Fourier Transform (STFT) [6] of a signal is used in some of the approaches, which basically calculated Fourier coefficients of the signal in temporal windows.

3.1.2 Continuous Wavelet Transform

The continuous wavelet transform is a method of analyzing the frequency components of a signal at specific time intervals. The advantage that CWT has over STFT is that it solves the problem of trade-off between frequency resolution and time resolution. When performing an STFT on a signal, one has to choose the window length for dividing the signal into sub-signals and performing DFT on each window, meaning that the larger the window size is set, a higher frequency resolution (the frequency components are better explained for the signal as a whole) and a lower time resolution (the changes of frequencies across time are not explained well) is obtained. The opposite holds as well: STFT with a smaller window size has higher time resolution but lower frequency resolution. CWT solves this problem of trade-off by representing the signal at different frequency scales, larger scale corresponding to lower frequencies, and lower scales to higher ones. At smaller scales, the signal is divided into smaller time windows, and lower frequency information is extracted, resulting in higher temporal resolution but lower frequency resolution. At larger scales, the signal is divided into larger time intervals, and higher frequency information is processed, resulting in higher frequency resolution but lower temporal resolution.

CWT makes use of wave-like functions called wavelets, and, at each step of the algorithm, the original signal is convolved by the wavelet function for deriving the corresponding frequency-domain value. The requirements for a function $f(t)$ to be considered a wavelet function as follows (complex wavelets are not considered in this paper, the following conditions relate to the real-valued wavelet qualifications only) [8]:

$$E = \int_{-\infty}^{\infty} |f(t)|^2 dt < \infty, \text{where } E \text{ is termed as the energy of } f,$$

$$\int_{0}^{\infty} \frac{|F(k)|^2}{k} dk < \infty, \text{where } F(k) \text{ is the Fourier transform of } f.$$

The most commonly used wavelet functions are Gaussian wave, Mexican hat, Haar and Morlet [8], the latter of which we utilized in speech signal processing (visualized in Fig. 1.)

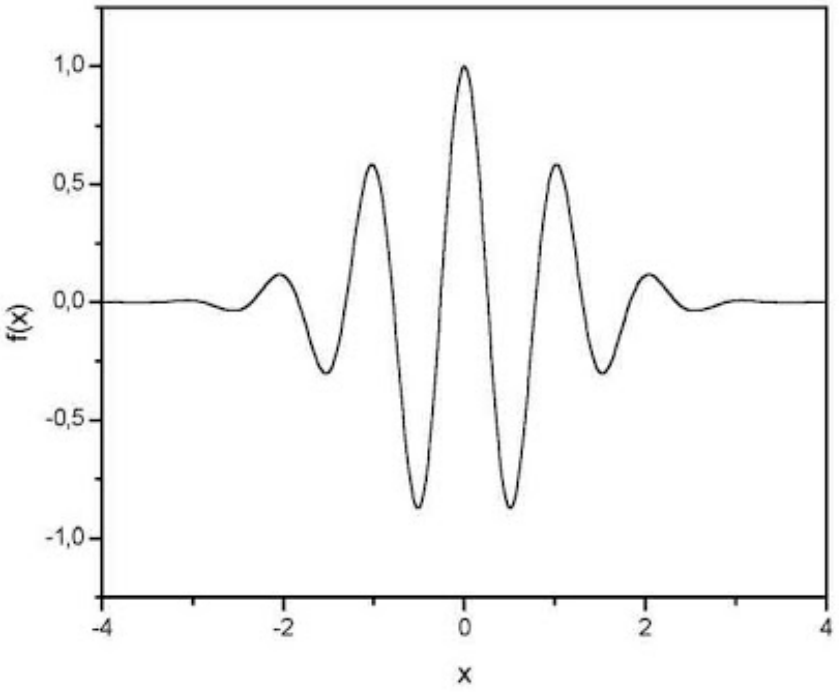

Fig. 1. Morlet wavelet function.

After choosing the wavelet function $\Phi(t)$, the CWT of the signal $x(t)$, denoted as $T(a, b)$, is computed as follows:

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \Phi \left(\frac{t-b}{a} \right) dt,$$

where a is the scale at which the signal is processed, and b indicates the time interval at which the signal is convoluted with the wavelet function. An example of a heatmap resulting from CWT is visualized in Fig. 2.

3.1.3 Mel Frequency Cepstral Coefficients

Another representation of audio signals that our methods use are Mel-frequency ceptstral coefficients (MFCC). MFCCs represent a temporal signal by cepstral energy coefficients at specific time intervals. The motivation of using MFCCs is to represent a signal by features that replicate the perception of audio signal by a human ear. Such representation is obtained by processing the signal with cepstral filters across frequency scales, the length of which is directly proportional to the scale of the frequency [9].

The resulting MFCC representation of a signal is given as a function $P_i(k)$, the output of which is the value of k-th cepstral coefficient at i-th temporal frame index. An example of an extracted MFCC feature is demonstrated in Fig. 3.

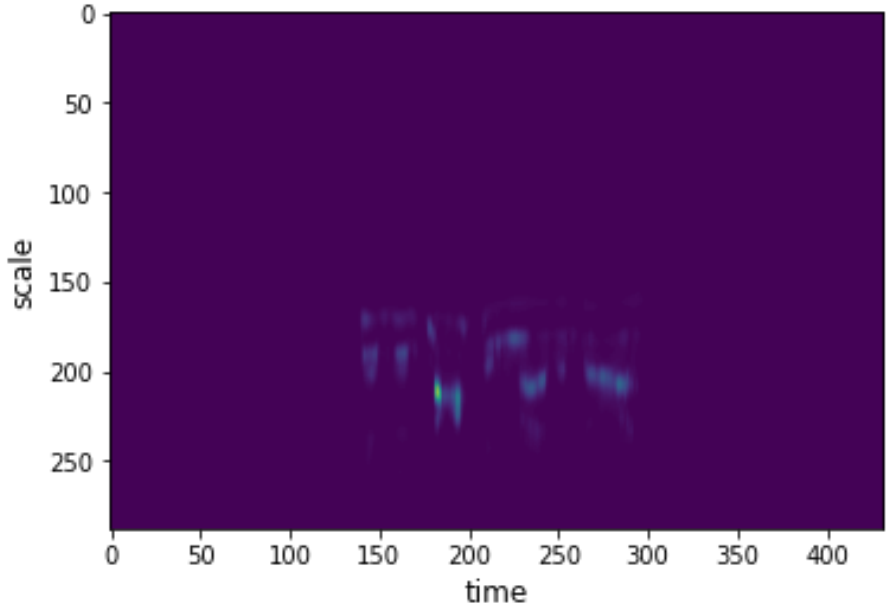

Fig. 2. Sample CWT heatmap of an audio signal.

3.2 Technical Details

For the extraction of audio signal features described in section 3.1, we use the "librosa" library for python [10]. "Pytorch" was used as a deep learning library for training our models [11]. The architectural details of each model are described in their respective sections.

3.3 Recognition Approaches

Having the labeled audio signals and their feature representations from CWT and MFCC, the next step is designing a method for emotion recognition from those signals. Following [12], the approach that this work relies on is using a sentiment-arousal space of emotions, which is depicted in Fig. 4. The idea is to come up with an intuitive 2-dimensional organization space of emotions by defining 2 axes: the arousal axis, and the positivity axis. By assigning these 2 values to every emotion label, we come up with an intuitive organization of emotion values in this space, as demonstrated in Fig. 4.

Having the sentiment-arousal space allows us to come up with different emotion recognition approaches, such as defining each quadrant of the 2D space as a classification label (i.e., whether the emotion is active-positive, active-negative, passive-positive, or passive-negative), or viewing the sentiment-arousal space as a continuous one, and solving the recognition task as a regression problem. In the upcoming sections, we show each of such approaches used along with the corresponding extracted features and the neural network architecture.

To the best of our knowledge, our proposed methods are the first try on tackling the problem in the specified setups. An exception is the setup of classification in sentiment-arousal space using CWT features and CNNs, where we compare to a method that has some of its aspects of setup shared with ours.

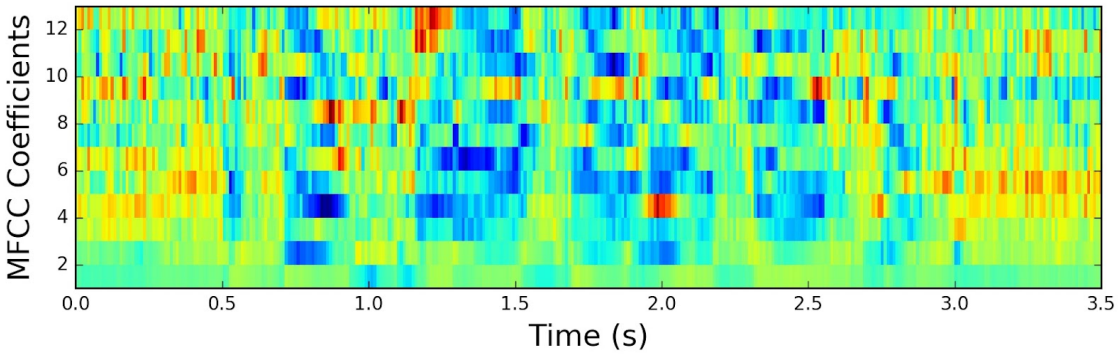

Fig. 3. Sample MFCC representation of a voice recording signal.

3.4 Architectures and Results

3.4.1 Mapping Emotion to Continuous Sentiment-Arousal Space

Table 3: Mapping of emotion values in sentiment-arousal continuous space

Emotion	Sentiment-Arousal Coordinates
Neutral	[0,0]
Calm	[0.25,-1]
Sad	[-0.75,-0.5]
Happy	[1,0.75]
Angry	[-0.75,1]
Fear	[-1,0.25]
Disgust	[-0.25,0.25]
Surprise	[0.25,1]
U.	1

An interesting approach that we can take towards the voice emotion recognition task is using the sentiment-arousal dimensions for defining a continuous space of emotion values, and solving a regression problem of emotion prediction. Specifically, for each emotion label coming from the datasets, we define sentiment and arousal values, as described in Tab. 3, which results in the organization of emotion values in a continuous sentiment-arousal space. Thus, the objective of the problem can be defined as:

$$L(\hat{y}) = \frac{1}{2}(\hat{y} - y)^T(\hat{y} - y) + \lambda \sum_{w \in W} w^2,$$

where $\hat{y}$ is the predicted point in the continuous space, y is the point in the 2D space corresponding to the ground-truth emotion label. W is the set of the trainable parameters, and thus the last term serves as a regularization to the optimization problem.

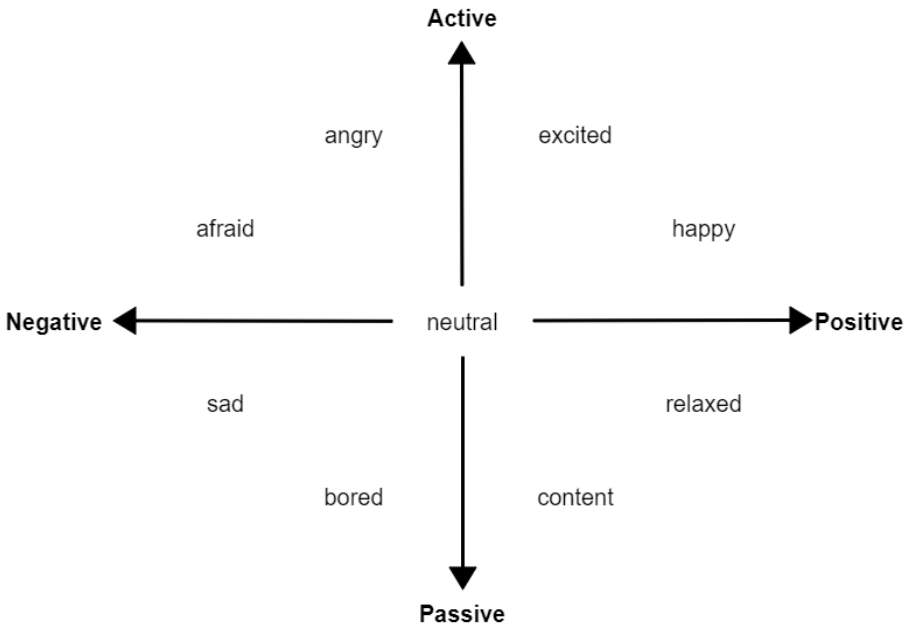

Fig. 4. The dimensions of the sentiment-arousal space, and how different emotions are organized in the space.

For solving the resulting regression task, we utilize MFCC features of audio recordings as inputs. We use CNN architecture for the model, which is shown in Fig. 5. Average pooling of size (2x2) is used for downsampling between the layers. The last layer is a fully connected layer that maps the flattened output of convolutional layers to the 2-dimensional sentiment-arousal space. Each layer has 32 output channels. The first layer has kernels of size (10x3), which is followed by a layer with (5x5) and a layer with (3x3) kernel sizes. Between layers, leaky relu activation function was used.

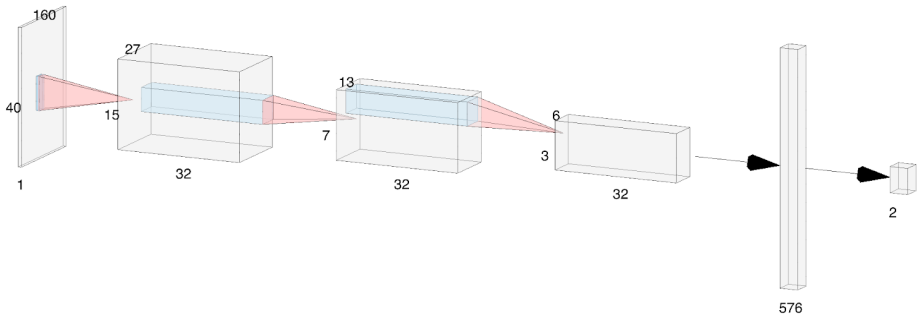

Fig. 5. CNN architecture used for the continuous emotion recognition task.

Fig. 6 shows the output of the model on voice recordings with the corresponding emotion labels. In the majority of cases, the network correctly identifies both the sentiment and

arousal of speech. It rarely fails to identify both components and it can at least identify the arousal of the speech. One of the shortcomings we see is that the significant proportion of the recordings with a "happy" label were identified as negative by the network. On the contrary, fear, disgust, anger and sadness were correctly positioned in the space. This, as also pointed out in the previous sections, shows us that the network is struggling to determine the positivity, but is good at differentiating between active and passive emotions.

Fig. 6. Performance of the CNN model on the continuous emotion recognition task.

3.4.2 Classification Using Sentiment-Arousal Space: LSTM with MFCCs

First, we solve a classification problem defined by the quadrants of the sentiment-arousal space. We use the extracted MFCC features as our input, and, viewing MFCC's as temporal signals, we use LSTMs [13] as our neural network architecture. Only the first 40 cepstral coefficients were considered. The datasets used were RAVDESS and TESS datasets (in some scenarios, only RAVDESS was considered.) For all classification models, for a single audio recording, given its ground-truth label values $\{y_1, y_2, ..., y_n\}$ and the estimated label values $\{\hat{y_1}, ..., \hat{y_n}\}$, the objective function is:

$$L = -\sum_i y_i \log(\hat{y_i}) + \lambda \sum_{w \in W} w^2,$$

where W is the set of all trainable weights.

There were 4 scenarios of splitting the dataset into train and test subsets: 1. 10% testing and 90% training (standard), 2. all the recordings of the first 2 actors as the test dataset and the rest as the train 3. all the recordings of the first 3 actors as the test dataset and the rest as the train, 4. all the recordings of the first 4 actor as the test dataset and the rest as the train. The architecture of LSTM model depicted in Fig. 7 was used for all scenarios. A dropout layer with probability p=0.3 was used.

The results of the experiment are summarized in Tab.4.

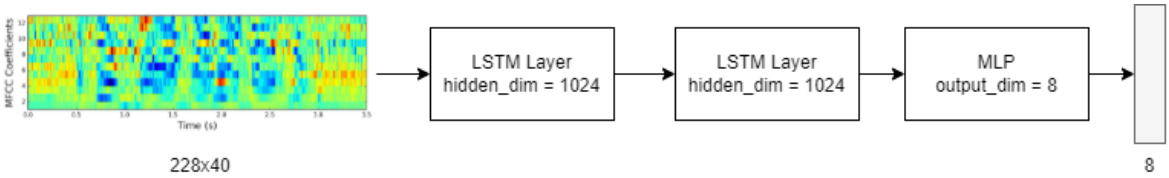

Fig. 7. The architecture of the trained LSTM model.

Table 4: Classification Results of LSTM model on different scenarios

Zones	Data separation	Datasets	Train acc.	Test acc.	AUC
4 zones	standard	RAVDESS	96.30%	67.36%	–
4 zones	2 new actors	RAVDESS	96.13%	74.16%	–
Arousal zones	standard	RAVDESS	97.76%	87.14%	0.91
Arousal zones	2 new actors	RAVDESS	99.54%	90.83%	0.94
Arousal zones	3 new actors	RAVDESS+TESS	94.23%	86.11%	0.91
Arousal zones	4 new actors	RAVDESS+TESS	98.40%	83.33%	0.87
Arousal zones	2 new actors	RAVDESS+TESS	95.63%	93.30%	0.97
Sentiment zones	standard	RAVDESS	98.30%	80.00%	0.81
Sentiment zones	2 new actors	RAVDESS+TESS	96.89%	84.14%	0.91
Sentiment zones	3 new actors	RAVDESS+TESS	93.69%	79.44%	0.84

From the results, we can see that the model managed to learn meaningful representations from the supervision signal. Since the same LSTM architecture gave performance for all classification scenarios, it indicates that the architecture is a good fit considering the datasets available. Also, the results indicate that the performance was good in classifying the arousal level of the speech, but classifying positivity is a bigger challenge for the model. This can be explained by the fact that MFCCs represent the energy amount in the signal in specific frequency or cepstral ranges, and, intuitively, larger amounts of energies correspond to higher arousal level. However, both negative and positive emotions can correspond to a high arousal level (i.e., surprised and angry), but it is harder to say how energy features can distinguish the positivity of a given speech.

3.4.3 Classification Using Sentiment-Arousal Space: CNN with CWT

The next experiment that we conducted is solving the problems of arousal level classification and positivity classification with CWT as inputs, and using CNN as the neural net architecture. Only RAVDESS dataset was considered in this experiment, and it was divided into a 10% test and 90% train datasets in both classification problems. Fig. 8. illustrates the architecture of CNN used for the classification tasks. Dropout with p=0.4 was used between each convolutional layer to prevent overfitting. Leaky ReLU was used as an activation function between layers and for preventing vanishing gradients. The results of the experiment

are summarized in Fig. 9 and in Tab. 5.

Table 5: Classification results of CNN model trained on CWT data

Zones	Train accuracy (%)	Test accuracy (%)	AUC
Arousal zones	98.70	83.76	0.84
Sentiment zones	87.76	75.71	0.77

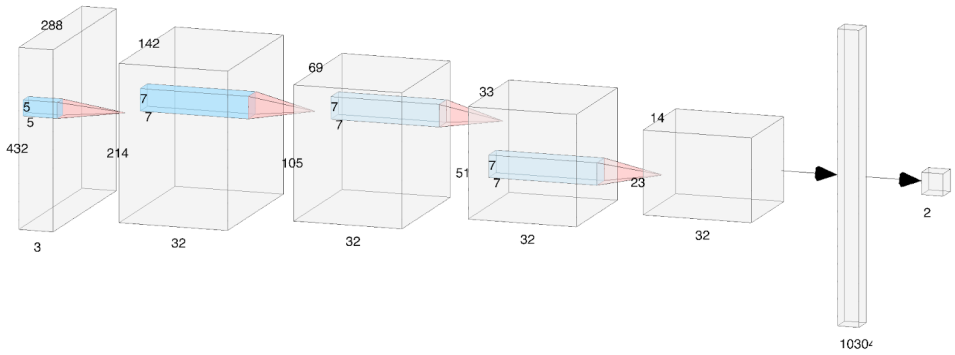

Fig. 8. CNN model architecture used for the CWT-based classification tasks.

As noted in the previous experiment, the models encounter the difficulty of classifying the positivity of the speeches.

[14] proposes methods for classifying arousal and sentiment in speech. They use the DEAP database [15], and their setup considers only "happiness", "sadness" and "cheerfulness" emotional labels. In their 2-label classification setting (high/low arousal; positive/negative sentiment), the arousal classification and sentiment classification accuracies are 61.23% and 92.19%, respectively, which are comparable results to our method.

4. Conclusion

This work proposes methods for solving voice emotion recognition tasks based on deep learning models. Audio signals were represented by features resulting from MFCC and CWT transforms. A pivotal component in the approaches is defining a 2D sentiment-arousal space, where the emotion values are organized in an intuitive way, allowing to define the recognition problem within this space either as a classification or a regression. The main challenge identified in all the proposed methods was the difficulty of recognizing the positivity aspect of the recordings, a possible explanation to which is the absence of such information in the features used, which mainly encode energies corresponding to frequency ranges. Overall, the results indicate that the models manage to learn features meaningful for the emotion recognition task. As one of the main challenges was the scarcity of labeled data, possible

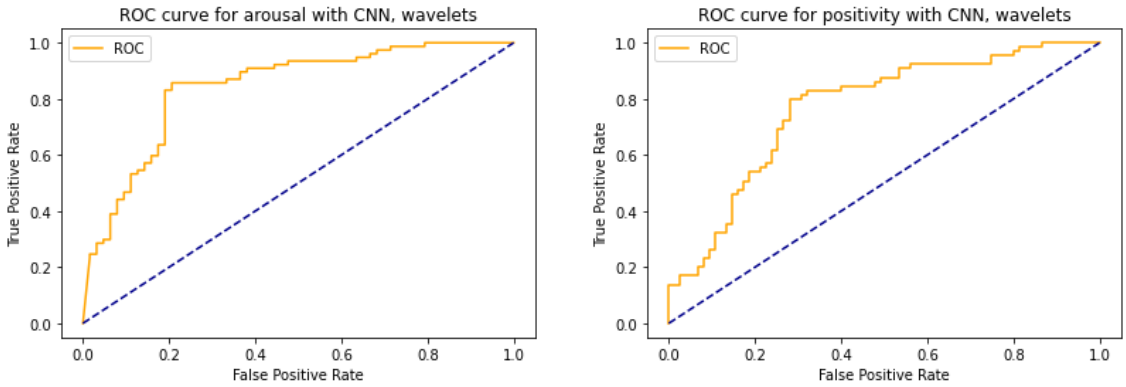

Fig. 9. ROC curves of the CNN classifier.

future directions include the use of data augmentations on voice recordings, as well as self-supervised approaches for learning semantic representations of the audio signals and fine-tuning those features for emotion recognition task, which doesn't require any labeled data.

References

[1] E. Mower, M. J.Mataric and S.Narayanan, "A framework for automatic human emotion classification using emotion profiles", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1057–1070, 2010.

[2] S. Glge, R. Bck and T. Ott,"Emotion recognition from speech using representation learning in extreme learning machines", *Proceedings of the 9th International Joint Conference on Computational Intelligence*, Funchal, Portugal, pp. 179–185, 2017.

[3] P. Jackson and S. Haq, "Surrey audio-visual expressed emotion (savee) database", University of Surrey: Guildford, UK. 2014.

[4] S.R.Livingstone, and F.A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English", *PLoS ONE* , vol. 13, no. 5, 2018.

[5] Pichora-Fuller, M. Kathleen and K. Dupuis, "Toronto emotional speech set (TESS)", Scholars Portal Dataverse, 2020. https://doi.org/10.5683/SP2/E8H2MF

[6] K. Grchenig, *Foundations of Time-Frequency Analysis*, First Edition. Birkhuser, Boston, MA, 2001.

[7] A. Kulkarni, M.F. Qureshi and M. JHA, "Discrete fourier transform: Approach to signal processing", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 03, pp. 12341–12348, 2014.

[8] P. S. Addison, *The Illustrated Wavelet Transform Handbook* , Second Edition. CRC Press, 2017.

[9] M. Sahidullah and G. Saha, "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition", *Speech Communication* , vol. 54, no. 4, pp. 543–565, 2012.

[10] B. McFee, A. Metsai, M. McVicar, S. Balke, C. Thom, C. Raffel, F. Zalkow, A. Malek, D. Kyungyun Lee, O. Nieto, D. Ellis, J. Mason, E. Battenberg, S. Seyfarth. (2022). librosa/librosa: 0.9.0 (0.9.0). Zenodo. https://doi.org/10.5281/zenodo.5996429

[11] A. Paszke, S.Gross, F.Massa, A. Lerer, J. Bradbury,G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library", *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, 2019.

[12] J. Posner, J.A. Russell and B.S. Peterson, "The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology", *Development and psychopathology*, vol. 17, no. 3, pp. 715–734, 2005.

[13] S. Hochreiter and J. Schmidhuber,"Long Short-Term Memory", *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] G. Garg and G. K. Verma, "Emotion recognition in valence-arousal space from multichannel EEG data and wavelet based deep learning framework", *Procedia Computer Science*, vol. 171, pp. 857–867, 2020.

[15] S. Koelstra, C. Muhl, M. Soleymani, J.S. Lee, A. Yazdani, T. Ebrahimi and I. Patras, "Deap: A database for emotion analysis; using physiological signals", *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

Խորը ուսուցման մեթոդներ ձայնագրությունների էմոցիայի գնահատման համար օգտագործելով տրամադրական կոորդինատային համակարգ

Նարեկ Տ. Թումանյան

Վեյցմանի գիտությունների համալսարան
e-mail: narek.tumanyan@weizmann.ac.il

Ամփոփում

Այս հոդվածում ներկայացվում են խորը ուսուցման վրա հիմնված մոտեցումներ՝ ձայնագրությունների էմոցիայի գնահատման խնդրի համար: Առաջադրված մոտեցումների բանալի բաղադրիչ է հանդիսանում էմոցիաների դասերի ներկայացումը տրամադրական երկչափ կոորդինատային համակարգում, որտեղ աքսիսների չափման միավոր են հանդիսանում էմոցիայի դրական/բացասական լինելը և ակտիվ/պասիվ լինելը, ինչպես նաև այդ ներկայացման օգտագործումը ուսուցման վերահսկման մեջ: Աուդիո ազդանշանները մշակելու համար օգտագործվել են ձայնագրությունների հաճախական տվյալներ: Որպես խորը ուսուցման մոդելներ, առաջադրված մեթոդներում օգտագործվում են լրիվ փաթույթային նեյրոնային ցանցեր (CNN) և երկար կարճաժամկետ հիշողություն (LSTM): Ներկայացվում են տարբեր էմոցիայի գնահատման մոտեցումներ և վերլուծվում են արդյունքները:

Բանալի բառեր՝ ձայնագրության էմոցիայի գնահատում, տրամադրական կոորդինատային համակարգ, հաճախական հատկանիշներ, խոսքի տրամադրության դասակարգում:

Глубокое обучение для распознавания эмоций в записях голоса с использованием валентно-возбужденного пространства

Нарек Т. Туманян

Институт Вейцмана, Израиль
e-mail: narek.tumanyan@weizmann.ac.il

Аннотация

В этой статье представлены основанные на глубоком обучении подходы к задаче распознавания эмоций в записях голоса. Ключевым компонентом этих методов является представление категорий эмоций в валентно-возбужденном пространстве, и использование этого пространства в качестве обучающего сигнала. Наш метод использует вейвлетные и кепстральные признаки для эффективного представления аудиосигнала. Для задачи распознавания были использованы сверточные нейронные сети (CNN) и сети долгой краткосрочной памяти (LSTM). Архитектура выбиралась в зависимости от того, каким образом был представлен сигнал - в пространственном или временном виде. Были использованы различные подходы к задаче распознавания, и были проанализированы результаты.

Ключевые слова: распознавание эмоций в голосе, валентно-возбужденное пространство, кепстральные признаки, классификация настроения голоса.

Mathematical Problems of Computer Science 56, 48–55, 2021.
doi: 10.51408/1963-0078

UDC 004.056.5

Differential Privacy in Practice: Use Cases

Karen A. Mastoyan

Gavar State University
e-mail: kmastoyan@yandex.com

Abstract

The problem of ensuring privacy is relevant in connection with the development of big data technologies. One of the modern and most promising methods of privacy protection is the differential privacy. In this paper the differential privacy applications developed by big companies are investigated. The libraries' capabilities and tools of Google, IBM, as well as packages in R are analyzed. The differential privacy process for data collected from users implemented by Apple is studied.
Keywords: Big data, Differential privacy, R environment.
Article info: Received 16 September 2021; accepted 3 November 2021.

1. Introduction

Big Data is an actual research topic because it provides new opportunities with data analysis for businesses and organizations to improve the decision making power. Various large companies, such as Facebook, Apple, Amazon and Google infiltrate users' personal lives and social interactions to accumulate huge databases at any time, which violates people's privacy. Along with the growth of data, it is necessary to develop new methods and means that will allow people to remain confidential.

A number of research articles are devoted to the study of Big data privacy, in the surveys [1] - [3] one can find detailed information and a full list of publications. In the role of information theory in the field of big data privacy is surveyed in [4].

There are various methods of confidentiality that allow for large-scale data analysis, statistical analysis, data (text) excavation, etc., while ensuring the privacy of individual participants. The most reliable approach is the Differential Privacy (DP).

DP is a modern approach to cyber security, where proponents argue that personal data is much better protected than traditional methods. DP is a strict mathematical definition of privacy [5], [6]. In the simplest terms, consider an algorithm that analyzes a dataset and calculates its statistics. It is said that such an algorithm is differentially private, if looking at the output, one cannot say whether anyone's data was included in the original database or not. In other words, the guarantee of a differential algorithm is that its behavior is unlikely to change when an individual joins or leaves a database. Anything that an algorithm can

retrieve in a database that contains some individual information is almost as likely to come from a database without that individual information.

Most importantly, this guarantee is reserved for any individual and any dataset. Therefore, no matter how strange someone's details are, and no matter how much someone's details are in the database, the guarantee of differential confidentiality is still maintained. This provides a formal guarantee that individual-level information about database participants will not be leaked.

DP protects an individual's privacy by adding a few random noises to the database when analyzing the data [7]. Simply put, identifying personal information based on the results of the analysis by presenting noise will not work. However, after adding noise, the result of the analysis turns into an approximation, not an accurate result, which would be obtained only if it were conducted on a real database. In addition, it is possible that if the differential private analysis is performed several times, it may yield different results each time the randomness of the noise is presented in the databases.

In this article we discuss the application of DP by big companies such as Apple, Google, IBM. We study the libraries developed by Google, IBM and a package for R developed by Benjamin I. P. Rubinstein, called Brubinstein's diffpriv package, and analyze the capabilities and tools in them.

2. Usage of differential privacy

It is worth noting that DP works better on larger databases. The reason is that as the number of individuals in a database increases, so does the impact of any individual on a given aggregate statistic. DP can be applied to everything from warranty systems and social networks to deployment-based services. Example:

- Apple employs DP to accumulate anonymous usage insights from devices like iPhones, iPads and Mac.
- Amazon uses DP to access user's personalized shopping preferences while covering sensitive information regarding their past purchases.
- Facebook uses it to gather behavioral data for target advertising campaigns without defying any nation's privacy policy.
- There are various variants of differentially private algorithms employed in machine learning, game theory and economic mechanism design, statistical estimation, and many more.

2.1 Apple

Apple has mastered and developed a technique known in academia as local differential privacy to do something very interesting: gain insight into what many Apple users are doing while helping to protect the privacy of individual users. It's a technique that allows Apple to learn about the community of users without knowing about individuals in the community [8]. DP transforms the information shared with Apple before it ever leaves the user's device, so that Apple can never reproduce the actual data. The DP technology used by Apple is rooted in the idea that statistical noise, which is a bit biased, can disguise a user's personal information

before it is shared with Apple. If many people share the same data, the added noise may on average exceed a large number of data points, and Apple can see that meaningful information is emerging. DP is used as the first step in a data analysis system that includes strong privacy protection at every stage. The system is opt-in and designed to ensure transparency to the user. The first step is to privatize information using local differential privacy on the user's device. The purpose of the privatization is to ensure that Apple's servers do not receive clear data. Device specifications are removed from the data and transmitted to Apple via an encrypted channel. Apple Analytics system ingests differential private contributions, droping IP addresses and other metadata. The final step is consolidation, where customized protocols are developed to calculate relevant statistics, and the consolidated statistics are then shared with Apple's respective teams. Both ingestion and consolidation phases take place in a restricted environment, so even privatized data is not widely available to Apple employees. Apple's implementation of DP includes the idea of a perdonation privacy budget (measured by parameter ϵ), and imposes a strict limit on the amount of data transmitted by a user to maintain their privacy. The fact is that the slightly biased noise used in DP tends to outperform a large number of investments on average, which theoretically allows us to determine user activity information over a large number of views per user (although it is important to note that Apple does not associate any characteristic with information collected through DP). Apple uses local differential privacy to help protect the privacy of users' activities over a period of time, while gaining insight that improves intelligence and usability of features such as:

- QuickType suggestions,
- Emoji suggestions,
- Lookup Hints,
- Safari Energy Draining Domains,
- Safari Autoplay Intent Detection (macOS High Sierra),
- Safari Crashing Domains (iOS 11),
- Health Type Usage (iOS 10.2).

For each feature, Apple seeks to reduce the privacy budget while collecting enough data for Apple to improve the features. Apple stores the collected data for a maximum of three months. The sent data does not include any identifiers and the IP addresses are not stored. For Lookup Hints, Apple uses a privacy budget with $\epsilon = 4$ and limits the sending of user data to twice a day. For Emoji, Apple uses a privacy budget with $\epsilon = 4$ and requires a one-time daily data submission. For QuickType, Apple uses a privacy budget with $\epsilon = 8$ and collects data to twice a day. For Health Type Usage, Apple uses a privacy budget of $\epsilon = 2$ and limits the sending of user data to once a day. The submitted data does not include the health information itself, but what types of health data are edited by the users. For Safari, Apple limits the transfer of user data twice a day. For Safari domains, which are known to cause high power consumption or crashes, Apple uses a single privacy budget with $\epsilon = 4$. For Safari Auto-play intentional detection, Apple uses a privacy budget of $\epsilon = 8$.

Apple uses the Count Mean Sketch technique for DP, with which the initial information that is processed for sharing with Apple is encrypted using a number of hash functions, making easy the data representation in different sizes of a fixed matrix.

The data is encrypted using the SHA-256 account variants, followed by the privatization step, and then written to a chart matrix, the values of which originate from zero.

The noise injection step works as follows: after encoding as a vector function, each coordinate of the vector is then bent (written as an incorrect value) with a probability of

$$1/(1 + e^{\epsilon/2}),$$

where ϵ is the privacy parameter. This ensures that the analysis of the collected data cannot distinguish real values from deviated values, helping to ensure the confidentiality of the shared information. To stay within the privacy budget, the Apple OS does not send the entire chart matrix to the server, but only a random array of matrices. When the information encoded in the graphical matrix is sent to Apple, the Apple server presents the responses of all the devices that share the information and subtracts the average value for each element of the array. Although each presentation contains a lot of random elements, the average value of a large number of presentations gives Apple meaningful data.

The Hadamard Count Mean-based Sketch technology uses the noise injection method, which is similar to the method used in the Count Mean Sketch method, but with one important difference. It uses a type of mathematical operation called converting the Hadamard base to hashed encoding before performing the privatization step. Also, it only sends 1 bit randomly instead of the whole series, as in the Count Mean Sketch technique. This reduces communication costs by 1 bit due to some accuracy.

For each feature, Apple seeks to make the privacy budget small while still collecting enough data to enable Apple to improve the features. Apple retains the collected data for a maximum of three months. The donations do not include any identifier, and IP addresses are not stored.

2.2 Google

Google DP repository contains libraries to generate ϵ- and (ϵ, δ)-differentially private statistics over datasets. It contains the following tools.

- Privacy on Beam is an end-to-end DP framework built on top of Apache Beam. It is intended to be easy to use, even by non-experts.
- Three "DP building block" libraries in C++, Go, and Java implement basic noise addition primitives and differentially private aggregations. Privacy on Beam is implemented using these libraries.
- A stochastic tester is used to help catch regressions that could make the DP property no longer hold.
- A DP accounting library is used for tracking privacy budget.
- Google DP repository includes a command line interface for running differentially private SQL queries with ZetaSQL. You can use the Privacy on Beam laboratory to generate differential private data [9].

Currently, the DP building block libraries support the following algorithms:

Table 1: Google DP library algorithms [9].

Algorithm	C++	Go	Java
Laplace mechanism	Supported	Supported	Supported
Gaussian mechanism	Supported	Supported	Supported
Count	Supported	Supported	Supported
Sum	Supported	Supported	Supported
Mean	Supported	Supported	Supported
Variance	Supported	Supported	Planned
Standard deviation	Supported	Supported	Planned
Quantiles	Supported	Supported	Supported
Automatic bounds approximation	Supported	Planned	Planned
Truncated geometric thresholding	Supported	Supported	Supported
Laplace thresholding	Supported	Supported	Supported
Gaussian thresholding	Planned	Supported	Supported

Implementations of the Laplace and Gaussian mechanism use secure noise generation. These mechanisms can be used to perform computations that aren't covered by the algorithms implemented in our libraries. The DP building block libraries are suitable for research, experimental or production use cases, while the other tools are currently experimental and subject to change [10].

2.3 IBM DP library

IBM differential-privacy-library is comprised of four major components:

1. Mechanisms. These are the building blocks of DP, and are used in all models that implement DP. Mechanisms have little or no default settings, and are intended for use by experts implementing their own models. They can, however, be used outside of models for separate investigations, etc.

2. Models. This module includes machine learning models with DP. Diffprivlib currently has models for clustering, classification, regression, dimensionality reduction and preprocessing.

3. Tools. Diffprivlib comes with a number of generic tools for differentially private data analysis. This includes differentially private histograms, following the same format as Numpy's histogram function.

4. Accountant. The BudgetAccountant class can be used to track the privacy budget and calculate the total privacy loss using advanced composition techniques [11].

2.4 Brubinstein‘s diffpriv package in R (library)

Brubinstein‘s diffpriv R package implements generic mechanisms for DP, along with sensitivity sampler that replaces exact sensitivity bounds with empirical estimates. As a result, diffpriv privatizes a wide range of procedures under random DP, automatically, without

mathematical analysis and in many cases achieving high utility. Diffpriv is available from [12] under an open-source license. Brubinstein's diffpriv package makes privacy-aware data science in R easy. It implements the formal framework of DP: differentially-private mechanisms can safely release computed statistics, models fit, or arbitrary structures derived on privacy-sensitive data to untrusted third parties. Due to the worst-case nature of the framework, mechanism development typically requires involved theoretical analysis. diffpriv offers a turn-key approach to DP by automating this process with sensitivity sampling in place of theoretical sensitivity analysis [13].

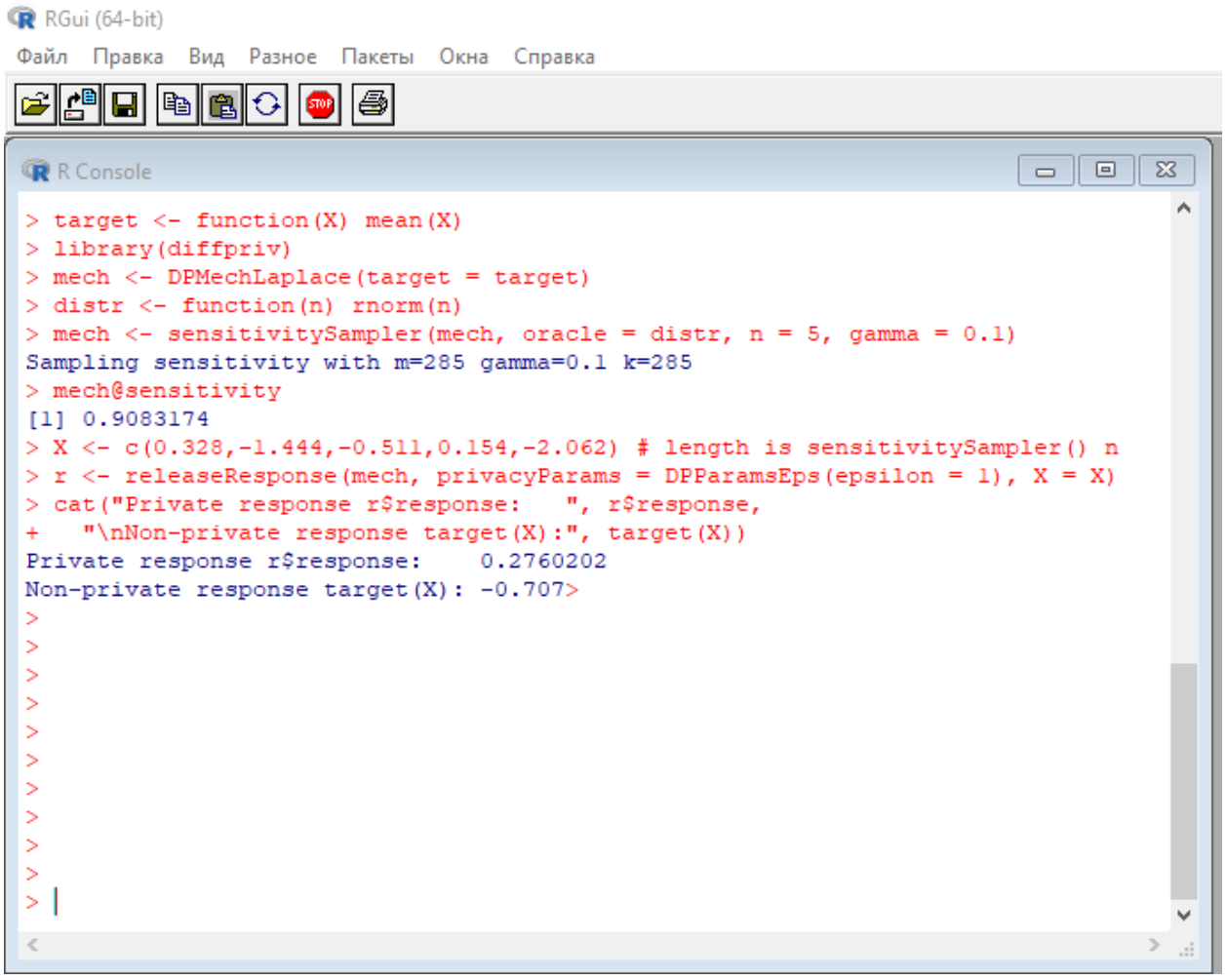

Fig. 1. Privately releasing a simple target function of privacy-sensitive input data X.

3. Conclusion

We investigated the implementation of DP by Apple, Google, IBM, as well as a package developed by Benjamin I. P. Rubinstein for R. As a result we can mention that Apple actively uses differential privacy in its systems, but does not provide open source libraries. The Google library is convenient for use with scripts written in C ++, Go, Java programming languages and allows the implementation of a number of algorithms, such as Laplace mechanism, Gaussian mechanism, Variance, Standard deviation, etc. IBM differential privacy library is written in python language and is more convenient to use in a python environment.

Brubinstein's diffpriv allows the implementation of the differential privacy process in R, which is open for various environments and algorithms. This research is useful for developing new DP applications and libraries.

References

[1] A. Mehmood, I. Natgunanathan, Y. Xiang, G. Hua and S. Guo, "Protection of Big Data privacy", *IEEE Access*, vol. 4, pp. 1821–1834, 2016, doi: 10.1109/ACCESS.2016.2558446.

[2] L. Xu, C. Jiang, J. Wang, J. Yuan and Y. Ren, "Information security in Big Data: Privacy and data mining" *IEEE Access*, vol. 2, pp. 1149–1176, 2014, doi: 10.1109/ACCESS.2014.2362522.

[3] S. Yu, "Big Privacy: Challenges and opportunities of privacy study in the age of Big Data", *IEEE Acces*, vol. 4, pp. 2751–2763, 2016, doi: 10.1109/ACCESS.2016.2577036.

[4] M. Haroutunian and K. Mastoyan, "The role of information theory in the field of Big Data privacy, *Mathematical Problems of Computer Science*, vol. 55, pp. 45 - 53, 2021.

[5] C. Dwork, M. Bugliesi, B. Preneel, V. Sassone, I. Wegener (eds) Automata, "Differential Privacy", *Languages and Programming. ICALP*, Lecture Notes in Computer Science, vol 4052. Springer, Berlin, Heidelberg, 2006. https://doi.org/10.1007/11787006

[6] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy", *Foundations and Trends in Theoretical Computer Science*, vol. 9, no.3-4, pp. 211407, 2014.

[7] K. M. P. Shrivastva, M. A. Rizvi and S. Singh, "Big Data privacy based on differential privacy a hope for Big Data," *Proc. Intern. Conf. on Computational Intelligence and Communication Networks*, Bhopal, India, pp. 776–781, 2014. doi: 10.1109/CICN.2014.167.

[8] Differential Privacy Team, Apple, Learning with Privacy at Scale, [Online]. Available: https://docs-assets.developer.apple.com/ml-research/papers/learning-with-privacy-at-scale.pdf

[9] End-to-end differential privacy solution, [Online]. Available: https://github.com/google/differential-privacy/tree/main/privacy-on-beam

[10] Google Developers, Google, Enabling developers and organizations to use differential privacy, [Online]. Available: https://developers.googleblog.com/2019/09/enabling-developers-and-organizations.html

[11] Naoise Holohan, Stefano Braghin, Pol Mac Aonghusa and Killian Levacher. Diffprivlib: The IBM Dierential Privacy Library, 2019 https://arxiv.org/pdf/1907.02444.pdf

[12] Differential privacy package using R, [Online]. Available: https://github.com/brubinstein/diffpriv

[13] B. I. P. Rubinstein and A. Francesco, "diffpriv: An R package for easy differential privacy", *Journal of Machine Learning Research*, vol. 18, pp. 1-5, 2017.

Դիֆերենցիալ գաղտնիություն գործնականում. կիրառման դեպքեր

Կարեն Ա. Մաստոյան

Գավառի պետական համալսարան
e-mail: kmastoyan@yandex.com

Ամփոփում

Գաղտնիության ապահովման խնդիրն արդիական է մեծ տվյալների տեխնոլոգիաների զարգացման հետ կապված: Գաղտնիության պաշտպանության ժամանակակից և ամենախոստումնալից մեթոդներից մեկը դիֆերենցիալ գաղտնիությունն է: Այս հոդվածում ուսումնասիրվում են խոշոր ընկերությունների կողմից մշակված գաղտնիության իրականացման տարբեր հավելվածներ: Վերլուծվում են Google-ի, IBM-ի գրադարանների և գործիքների հնարավորությունները, ինչպես նաև՝ դիֆերենցիալ գաղտնիությունը R փաթեթում: Ուսումնասիրվում է օգտատերերից հավաքագրված տվյալների դիֆերենցիալ գաղտնիության գործընթացը՝ ներդրված Apple-ի կողմից: Այս հետազոտությունը օգտակար է գաղտնիության ապահովման նոր հավելվածներ և գրադարաններ մշակելու համար:

Բանալի բառեր՝ Մեծ տվյալներ, դիֆերենցիալ R միջավայր:

Дифференциальная конфиденциальность на практике: варианты применения

Карен А. Мастоян

Гаварский государственный университет
e-mail: kmastoyan@yandex.com

Аннотация

Проблема обеспечения конфиденциальности актуальна в связи с развитием технологий больших данных. Одним из современных и наиболее перспективных методов защиты конфиденциальности является дифференциальная конфиденциальность. В этой статье исследуются приложения дифференциальной конфиденциальности, разработанные крупными компаниями. Анализируются возможности библиотек и инструментов Google, IBM, а также пакетов в R. Изучается процесс дифференциальной конфиденциальности, реализованный Apple, для данных, собранных от пользователей. Это исследование полезно для разработки новых приложений и библиотек дифференциальной конфиденциальности.

Ключевые слова: Большие данные, дифференциальная конфиденциальность, среда R.

Mathematical Problems of Computer Science 56, 56–64, 2021.
doi:10.51408/1963-0079

UDC 519.688

Determining the Degree of Fuzzy Regularity of a String

Armen H. Kostanyan

IT Educational and Research Center
Yerevan State University
e-mail: armko@gmail.com

Abstract

The paper deals with the issue of determining the degree of fuzzy regularity of a crisp string. It is assumed that the concept of fuzzy regularity is formalized by a pattern given as a finite automaton with fuzzy properties of alphabet characters on transitions. Proceeding from this, we replace the problem of determining the degree of fuzzy regularity of a crisp string with the problem of determining the degree of belonging of such a string to the language of the corresponding automaton and propose an effective method for solving it using the dynamic programming approach.

The solution to the considered problem makes it possible to fuzzify the set of strings in a given alphabet based on a pattern defining fuzzy regularity. This work is a continuation of the author's previous works related to finding occurrences of a fuzzy pattern in the text. It may have applications in the field of pattern recognition, data clustering, bio-informatics, etc.

Keywords: Fuzzy string matching, Pattern recognition, Fuzzy automaton.
Article info: Received 4 October 2021; accepted 6 December 2021.
Acknowledgement: This work was supported by the Ministry of Education, Science, Culture and Sports of the Republic of Armenia, project 21T-1B326.

1. Introduction

This paper refers to the definition of a degree of fuzzy regularity of a crisp string in a given alphabet. To formalize the concept of fuzzy regularity, we use the highest degree of matching of such a string to the elements of a given set of periodical sequences of fuzzy properties of the alphabetic characters. We treat this set of sequences as a *matching pattern*.

As we noted, the sequences of fuzzy properties included in the pattern must have a certain periodicity. To achieve this goal, we propose to define the matching pattern as a finite automaton with fuzzy properties of alphabetic characters on transitions. As the pumping lemma states, any sufficiently large sequence of fuzzy properties accepted by such an automaton will be periodic with a period size not exceeding the number of states of

the automaton. Thus, in the concept we propose, the problem of determining the fuzzy regularity of a crisp string is defined as a problem of determining the degree of belonging such a string to the language of a fuzzy automaton.

It should be noted that the concept of a fuzzy automaton, which we use as a pattern for determining the regularity, is somewhat different from the one generally accepted in the literature. For example, in the concept of a fuzzy automaton used in [1] and [2], the automaton states and the alphabetic characters are assumed to be crisp, while the start and final states as well as the transitions of the automaton are assumed to be fuzzy. A review of the results on fuzzy automata in their various interpretations, as well as the fuzzy languages they accept (including automaton analysis and synthesis, minimization, closure properties, etc.) is given in [3].

The problem of determining the regularity of a string is widely used in *pattern recognition*, where it is often necessary to detect regularities in strings encoding images [4]. Such regularity is not always specified precisely, so we have to deal with a fuzzy regularity. The concept of fuzzy regularity can also be used in the area of fuzzy *data clustering* [5], which deals with grouping elements into fuzzy clusters, a particular case of which is fuzzy *sequence labeling* [6].

The investigation on fuzzy regularity of a string that we propose in this paper continues our previous research in the field of fuzzy string matching [7], [8], [9]. The problem of splitting a string into adjacent segments to best match the pattern, which is a sequence of fuzzy properties of substrings, was considered in [10], [11]. In this paper, we concretize the concept of a fuzzy property of a substring, defining it as the degree of belonging the substring to the language of a fuzzy automaton.

The paper is organized as follows.

Section 2 presents the concepts of a fuzzy symbol and a finite automaton over a set of fuzzy symbols, called a *fuzzy automaton*. Section 3 considers the problem of matching a crisp string with a pattern given by a fuzzy automaton and provides an efficient algorithm for determining the degree of matching based on the dynamic programming approach. Finally, the conclusion summarizes the obtained results.

2. Preliminaries

2.1 Fuzzy Symbols

Suppose that $(L, \leq, \otimes, 0, 1)$ is a finite linearly ordered set of measures with the smallest element 0, the largest element 1, and the monotonic accumulation operation $\otimes$ such that L is a commutative monoid with unit element 1 and zero element 0. That is, for all $a, b, c \in L$

$$a \otimes 0 = 0, a \otimes 1 = a, a \leq b \Rightarrow a \otimes c \leq b \otimes c.$$

In our further considerations, we will assume that the maximum over the empty set of L - values is equal to 0.

According to [12], the fuzzy subset X of the universal set U is defined by the membership function $\mu_X : U \to L$ that associates with each element $u \in U$ the value $\mu_X(u) \in L$, representing the degree of belonging u to X. A fuzzy subset X of U can be represented by the additive form

$$X = \sum_{u \in U} u/\mu_X(u).$$

We say that an element $u \in U$ certainly belongs to X if $\mu_X(u) = 1$, and it certainly does not belong to X if $\mu_X(u) = 0$. Conversely, if $0 < \mu_X(u) < 1$, then we say that u belongs to X with degree $\mu_X(u)$.

Given an alphabet Σ of characters, we define a *fuzzy symbol* α over Σ as a fuzzy subset of Σ. Given a character $x \in \Sigma$ and a fuzzy symbol α over Σ, we say that x matches α with degree $\mu_\alpha(x)$. This definition can be extended in the usual way to equal length sequences of characters and fuzzy symbols, respectively. That is, for a set of fuzzy symbols Ξ, $x = x_1...x_n \in \Sigma^*$ and $\omega = \omega_1...\omega_n \in \Xi^*$, we define the matching degree of x to ω as the L-value

$$\mu_\omega(x) = \begin{cases} \mu_{\omega_1}(x_1) \otimes ... \otimes \mu_{\omega_n}(x_n), & \text{if } x \neq \epsilon, \\ 1, & \text{if } x = \epsilon. \end{cases}$$

Example 1 . Suppose $\Sigma = \{1, 2, 3, 4, 5, 6, 7\}$ and the measures are rational numbers that belong to the segment $[0, 1]$ with the accumulation operation defined as multiplication. Define the fuzzy symbols S (*small*), M (*middle*) and L(*large*) to be the following fuzzy subsets of Σ:

- $S = 1/1 + 2/0.9 + 3/0.6 + 4/0.3 + 5/0.1 + 6/0 + 7/0$,
- $M = 1/0 + 2/0.25 + 3/0.75 + 4/1 + 5/0.75 + 6/0.25 + 7/0$,
- $L = 1/0 + 2/0 + 3/0.1 + 4/0.3 + 5/0.6 + 6/0.9 + 7/1$.

Let $x = 35634$, $\omega_1 = SMLSM$, $\omega_2 = MLMSL$. Then

- $\mu_{\omega_1}(x) = \mu_S(3) \otimes \mu_M(5) \otimes \mu_L(6) \otimes \mu_S(3) \otimes \mu_M(4) = \frac{3}{5} \cdot \frac{3}{4} \cdot \frac{9}{10} \cdot \frac{3}{5} \cdot 1 = \frac{243}{1000}$,
- $\mu_{\omega_2}(x) = \mu_M(3) \otimes \mu_L(5) \otimes \mu_M(6) \otimes \mu_S(3) \otimes \mu_L(4) = \frac{3}{4} \cdot \frac{3}{5} \cdot \frac{1}{4} \cdot \frac{3}{5} \cdot \frac{3}{10} = \frac{81}{4000}$.

2.2 Fuzzy Automaton

Given a finite set Ξ of fuzzy symbols over the alphabet Σ, we define a *fuzzy automaton* as a deterministic finite automaton using symbols from Ξ on transitions. That is, a fuzzy automaton is a 5-tuple $A = (Q, \Xi, \delta, q_{in}, F)$, where

- Q is the finite non-empty set of states,
- $\delta \subseteq Q \times \Xi \to Q$ is the transition function,
- $q_{in} \in Q$ is the initial state,
- $F \subseteq Q$ is the set of final states.

Let

$$h = p_0 \xrightarrow{\alpha_1} p_1 \xrightarrow{\alpha_2} ... \xrightarrow{\alpha_n} p_n,\ n \geq 0,$$

be a path leading from the state $p_0 \in Q$ to the state $p_n \in Q$ in the diagram of the fuzzy automaton A. We say that the path h generates a sequence $\alpha = \alpha_1...\alpha_n \in \Xi^*$ of fuzzy symbols. We define the language $L(A) \subseteq \Xi^*$ of the automaton A as the set of all sequences of fuzzy symbols generated by all paths leading from the initial state to a final state. For $k \geq 0$ we denote by $L(A)/k$ the set of all k-length strings in $L(A)$.

Given a string $x \in \Sigma^*$ and a fuzzy automaton A, we say that x matches A with degree $\mu_A(x)$, if

$$\mu_A(x) = max\{\mu_\omega(x) | \text{ for all } \omega \in L(A)/|x|\}$$

Example 2 . Let $x = 56364$, $|x| = 5$, $\Xi = \{S, M, L\}$, A is the fuzzy automaton in Fig. 1.

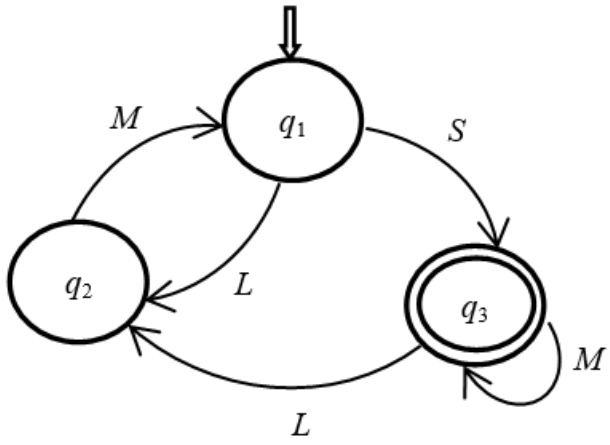

Fig. 1. A fuzzy automaton

Listed below are all the strings in $L(A)/5$:

$\omega_1 = SMMMM,\ \omega_2 = SMLMS,\ \omega_3 = SLMSM,\ \omega_4 = LMSMM,\ \omega_5 = LMLMS.$

Note that

$$\mu_A(x) = max\{\mu_{\omega_1}(x), \mu_{\omega_2}(x)\mu_{\omega_3}(x), \mu_{\omega_4}(x), \mu_{\omega_5}(x)\} = max\Big\{\mu_{SMMMM}(56364), \mu_{SMLMS}(56364),$$
$$\mu_{SLMSM}(56364), \mu_{LMSMM}(56364), \mu_{LMLMS}(56364)\Big\} = max\left\{\left(\frac{1}{10}\cdot\frac{1}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}\cdot 1\right),\right.$$
$$\left.\left(\frac{1}{10}\cdot\frac{1}{4}\cdot\frac{1}{10}\cdot\frac{1}{4}\cdot\frac{3}{10}\right), \left(\frac{1}{10}\cdot\frac{9}{10}\cdot\frac{3}{4}\cdot 0\cdot 1\right), \left(\frac{3}{5}\cdot\frac{1}{4}\cdot\frac{3}{5}\cdot\frac{1}{4}\cdot 1\right), \left(\frac{3}{5}\cdot\frac{1}{4}\cdot\frac{1}{10}\cdot\frac{1}{4}\cdot\frac{3}{10}\right)\right\} =$$
$$= max\left\{\frac{3}{640}, \frac{3}{16000}, 0, \frac{9}{400}, \frac{9}{8000}\right\} = \frac{9}{400}.$$

3. Calculation of Fuzzy Regularity

3.1 The Fuzzy Regularity Determination Problem

Let Σ be a finite alphabet of characters, Ξ be a finite set of fuzzy symbols over Σ, P be a fuzzy automaton over Ξ called a *fuzzy regularity pattern* (or, in short, a *regularity pattern*).

For a given $x \in \Sigma^*$, we define the *(x, P) - matching problem* as the problem of determining the value $\mu_P(x)$. We assume that this value represents the degree of fuzy regularity of the crisp string x according to the regularity pattern P.

3.2 Recursive Solution

Let $P = (Q, \Xi, \delta, q_{in}, F)$ be a regularity pattern, $q \in Q$.

Let us denote $P_q = (Q, \Xi, \delta, q, F)$ the regularity pattern obtained from P by replacing the initial state q_{in} with the state q. The value $\mu_{P_q}(x)$, representing the solution to the (x, P_q) - matching problem for the string x and the regularity pattern P_q, let us denote $\mu_P(q, x)$. In particular, we have that $\mu_P(q_{in}, x) = \mu_P(x)$.

Theorem 1: (Optimal substructure of the (x, P_q) - matching problem)

1. $x = \epsilon \Rightarrow [\text{If } \ q \in F \ \text{ then } \ \mu_P(q, x) = 1 \ \text{ else } \ \mu_P(q, x) = 0]$.
2. $x = ax' \Rightarrow [\mu_P(q, x) = max\{\mu_\alpha(a) \otimes \mu_P(q', x') | \text{ for all } \ \delta(q, \alpha) = q'\}]$.

Proof: The first statement is obvious.

The second statement follows from

$[x = ax'] \Rightarrow$

$$[\mu_P(q, x) = max\{\mu_\alpha(a) \otimes \mu_\omega(x') |\ \delta(q, \alpha) = q', \alpha \in \Xi, \omega \in L(P_{q'})] \Rightarrow$$
$$[\mu_P(q, x) = max\{\mu_\alpha(a) \otimes \mu_P(q', x') | \text{ for all } \delta(q, \alpha) = q'\}].$$

Theorem 1 implies the following recurrent equation for calculation $\mu_P(q, x)$:

$$\mu_P(q, x) = \begin{cases} (q \in F)?1 : 0, & \text{if} \quad x = \epsilon, \\ max\{\mu_\alpha(a) \otimes \mu_P(q', x') | \text{ for all } \delta(q, \alpha) = q'\}, & \text{if } x = ax'. \end{cases}$$

□

Direct calculation of the value of $\mu_P(q, x)$ using this formula will be inefficient due to overlapping subproblems. In order to get a more efficient solution, let us apply the dynamic programming approach.

3.3 Dynamic Programming Solution

Suppose Q is represented as an m-tuple $< q_1, ..., q_m >$, so that $q_{in} = q_1$. For $1 \leq i \leq m, 1 \leq j \leq n + 1$, let us denote $s[i, j] = \mu_P(q_i, x[j..n])$ (we assume that $x[n + 1..n] = \epsilon$).

The optimal substructure of the (x, P_q) - matching problem dictates the following recurrent equation for calculating $s[i, j]$:

$$s[i, j] = \begin{cases} q_i \in F)?1 : 0, & \text{if } j = n + 1, \\ max\{\mu_\alpha(x[j]) \otimes s[k, j + 1] | \text{ for all } \delta(q_i, \alpha) = q_k\}, & \text{if } 1 \leq j \leq n. \end{cases} \quad (1)$$

Note that the (x, P) - matching problem can be represented as the problem of determining the value $s[1, 1] = \mu_P(x)$.

For $q \in Q$, let us denote $out(q) = \{(\alpha, p) | \delta(q, \alpha) = p\ \}$, which is the set of pairs consisting of states and fuzzy symbols corresponding to transitions outgoing from q.

The algorithm in Fig. 2 presents the process of calculating the matrix $\{s[i, j]|, 1 \leq i \leq m, 1 \leq j \leq n + 1\}$ of matching degrees according to formula (1):

Algorithm 1: $Calculate - Matching - Degrees$

Input:
An n-length string x and an m-state regularity pattern P
Output:
The L-value matrix $s[1..m, 1..n+1]$ of matching degrees

```
for i = 1 to m do
    s[i, n+1] = ((q_i ∈ F)?1 : 0)
end
for j = n downto 1 do
    for i = 1 to m do
        max = 0
        for all (α, q_k) ∈ out(q_i)
        if μ_α(x[j]) ⊗ s[k, j+1] > max then
            max = μ_α(x[j]) ⊗ s[k, j+1]
        end
        s[i, j] = max
    end
end
return s
```

Fig. 2. Building the matrix of matching degrees

The solution to the (x, P) - matching problem, presented in Fig. 3, is simply reduced to extracting the value $s[1, 1]$ from the matrix s.

Algorithm 2: $Determine - Fuzzy - Regularity$

Input:
An n-length string x
An m-state regularity pattern P, where $Q =< q_1, ..., q_m >$ and $q_{in} = q_1$
Output:
The degree of matching x to P

```
s = Calculate − Matching − Degrees(x, P)
return s[1, 1]
```

Fig. 3. Determination of fuzzy regularity

Example 3 . The matrix s of matching degrees, constructed by the $Calculate-Matching-Degrees$ algorithm on the input $x = 56364$ and the regularity pattern in Fig. 1, is presented in Fig. 4.

	5	**6**	**3**	**6**	**4**	ϵ
q_1	$\mu_L(5) \cdot \frac{3}{80} = \frac{9}{400}$	$\mu_S(6) \cdot \frac{3}{16} = 0$	$\mu_S(3) \cdot \frac{1}{4} = \frac{3}{20}$	$\mu_S(6) \cdot 1 = 0$	$\mu_S(4) \cdot 1 = \frac{3}{10}$	0
q_2	*doesn't matter*	$\mu_M(6) \cdot \frac{3}{20} = \frac{3}{80}$	0	$\mu_M(6) \cdot \frac{3}{10} = \frac{3}{40}$	0	0
q_3	*doesn't matter*	$\mu_M(6) \cdot \frac{3}{16} = \frac{3}{64}$	$\mu_M(3) \cdot \frac{1}{4} = \frac{3}{16}$	$\mu_M(6) \cdot 1 = \frac{1}{4}$	$\mu_M(4) \cdot 1 = 1$	1

Fig. 4. Memoization results

In line with $Determine - Fuzzy - Regularity$ algorithm, the value $s[1,1] = \frac{9}{400}$ is the solution to the (x, P) - matching problem for the string $x = 56364$ and regularity pattern in Fig. 1, which is consistent with the result obtained in Example 2. According to our approach, this value determines the degree of fuzzy regularity of the crisp string x based on the pattern P.

3.4 Analysis

To estimate the complexity of the proposed solution to the (x, P)-matching problem, let us assume that there are k transitions in the graph of the m-state fuzzy automaton P, and that this graph is represented as an array $A[1..m]$ such that $A[i] = out(q_i), 1 \leq i \leq m$.

In this case, the construction of the previous column of the matrix s in lines 5-12 of the $Calculate - Matching - Degrees$ algorithm takes time $O(1 + k)$ and, consequently, the construction of the entire matrix s for a string of length n takes time $O(n(1 + k))$. As a result, the instruction in line 1 of the $Determine - Fuzzy - Regularity$ algorithm runs in $O(n(1 + k))$ time. The instruction in line 2 of this algorithm obviously runs in O(1) time, which makes the time complexity of the $Determine - Fuzzy - Regularity$ algorithm equal to $O(n(1 + k))$.

For an n-length string x and an m-state regularity pattern P, the algorithm uses $O(mn)$ extra memory to represent the matrix $s[1..m, 1..n + 1]$ of matching degrees.

4. Conclusion

The problem of determining the fuzzy regularity of a crisp string has been considered in this paper, where the concept of fuzzy regularity is formalized by means of a finite automaton with fuzzy properties of alphabet characters on transitions. Using the dynamic programming approach, we propose a solution to this problem with

$O(n(k + 1))$ time complexity, and

$O(mn)$ space complexity,

where n is the length of the input string; m and k are the number of states and the number of transitions of the automaton, respectively.

The proposed algorithm can be used in the field of pattern recognition, data clustering, DNA analysis, etc.

References

[1] Y. Cao and Y. Ezawa, "Non-determinstic fuzzy automata", Information Sciences, vol. 191, pp. 86-97, 2012.

[2] J. N. Mordeson and D.S. Malik, *Fuzzy Automata and Languages: Theory and Applications*, Chapman & Hall, CRC, Boca Raton, London 2002.

[3] R. K. Singh, A. Rani and M.K. Sachan, "Fuzzy automata: A quantitative review", *International Journal on Future Revolution in Computer Science & Communication Engineering*, vol 3, no. 7, pp. 11-17, 2017.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[5] Wang Zhen Zhou, "Image segmentation by combining the global and local properties", *Expert Systems with Applications*, vol. 87, pp. 30-40, 2017.

[6] Bezdek, James C, *Pattern Recognition with Fuzzy Objective Function Algorithms*, 1981.

[7] A. Kostanyan, Fuzzy string matching with finite automata, *Proc. on 2017 IEEE Conference CSIT-2017*, Yerevan, Armenia, pp. 25-29. IEEE Press, USA 2018.

[8] A. Kostanyan, Fuzzy string matching using prefix table, *Transactions of IIAP NAS RA, Mathematical Problems of Computer Science*, vol. 54, 2020, pp.116 -121.

[9] A. Kostanyan and A. Karapetyan, *String Matching in Case of Periodicity in the Pattern*, In: Dolinina O., Brovko A., Pechenkin V., Lvov A., Zhmud V., Kreinovich V. (eds) Recent Research in Control Engineering and Decision Making. ICIT 2019. Studies in Systems, Decision and Control, vol 199. Springer, 2019.

[10] A. Kostanyan and A. Harmandayan, Segmentation of string to match a fuzzy pattern, *In Proc. of Computer Science & Information Technologies (CSIT) Conference*, Yerevan, Armenia, September, pp. 17-19, 2019.

[11] A. Kostanyan and A. Harmandayan, Mapping a fuzzy pattern onto a string, *in Proc. on 2019 IEEE Conference "2019 Computer Science and Information Technologies (CSIT)*, Yerevan, Armenia, 23-27 Sep., 2019, IEEE Press, USA, pp. 5-8, 2019.

[12] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning-I, *Information Sciences*, vol. 8, pp. 199-249, 1975.

Տողի ոչ հստակ կանոնավորության աստիճանի որոշում

Արմեն Հ. Կոստանյան

ՏՏ կրթական և հետազոտական կենտրոն
Երևանի պետական համալսարան
e-mail: armhko@gmail.com

Ամփոփում

Հոդվածում դիտարկվում է տողի ոչ հստակ կանոնավորության չափի որոշման խնդիրը: Ենթադրվում է, որ ոչ հստակ կանոնավորության հասկացությունը ֆորմալիզացվում է վերջավոր ավտոմատի (շաբլոնի) միջոցով, որի անցումներին վերագրված են այբուբենի նշանների ոչ հստակ հատկություններ: Արդյունքում, հստակ տողի ոչ հստակ կանոնավորության չափի որոշման խնդիրը փոխարինվում է համապատասխան ավտոմատի լեզվին նշված տողի պատկանելիության չափի որոշման խնդրով, որի լուծման համար առաջարկվում է օգտագործել դինամիկ ծրագրավորման մեթոդը:

Նշված խնդրի լուծումը հնարավորություն է տալիս կառուցել տրված այբուբենի բառերի ոչ հստակ բազմություն՝ ոչ հստակ կանոնավորության որոշման շաբլոնի հիման վրա: Տվյալ աշխատանքը շարունակությունն է հեղինակի նախորդ աշխատանքների՝ տողում ոչ հստակ ենթատողի (շաբլոնի) որոնման ուղղությամբ: Այն կարող է կիրառվել շաբլոնների ճանաչման, տվյալների խմբավորման, բիոինֆորմատիկայի և այլ բնագավառներում:

Բանալի բառեր՝ նմուշի ոչ հստակ համադրում, շաբլոնների ճանաչում, ոչ հստակ ավտոմատ:

Определение степени нечеткой регулярности строки

Армен Г. Костанян

Образовательный и научный центр информационных технологий
Ереванский государственный университет
e-mail: armhko@gmail.com

Аннотация

В статье рассматривается задача определения степени нечеткой регулярности данной строки. Предполагается, что нечеткая регулярность формализуется посредством патерна, представленного в виде конечного автомата с нечеткими свойствами символов алфавита на переходах. В результате, задача определения степени нечеткой регулярности строки заменяется задачей определения степени ее принадлежности языку нечеткого автомата, для решения которой предлагается использовать метод динамического программирования.

Решение рассматриваемой задачи позволяет фаззифицировать множество слов в данном алфавите на основе паттерна определения нечеткой регулярности. Данная работа является продолжением ряда предыдущих работ автора по поиску нечеткого паттерна в строке. Она может иметь применения в таких областях, как распознавание паттернов, кластеризация данных, боинформатика, и т. д.

Ключевые слова: нечеткое сопоставление с образцом, распознавание паттернов, нечеткий автомат.

Mathematical Problems of Computer Science 56, 65–72, 2021.
doi:10.51408/1963-0080

UDC 510.6

On Proof Complexity of Some Type of Tautologies

Vahagn N. Altunyan and Garik V. Petrosyan

Yerevan State University
e-mail: altunyanv@gmail.com, garik.petrosyan.1@gmail.com

Abstract

In this paper, we investigate the proof complexities of a special type of tautologies, which are described as tautologies consisting of implications and literals. In particular, we prove that the proof of this kind of tautologies can be polynomially reduced to the proof of tautologies consisting of formulas that are described by sign-alternating trees.
Keywords: Frege systems, Tautology, Sign-alternating tree, Proof complexity.
Article info: Received 8 September 2021; accepted 18 November 2021.

1. Introduction

One of the most fundamental problems of the proof complexity theory is to find an efficient proof system for classical propositional calculus. There is a widespread understanding that polynomial-time computability is the correct mathematical model of feasible computation. According to the opinion, a truly effective system should have a polynomial-size $p(n)$ proof for every tautology of size n. In [1], Cook and Reckhow named such a system a super system. They showed that $NP = coNP$ iff there exists a super system. It is well known that many systems are not super. This question about the Frege system, the most natural calculi for propositional logic, is still open.

In many papers, some specific sets of tautologies are introduced, and it is shown that the question about polynomially bounded sizes for Frege-proofs of all tautologies is reduced to an analogous question for a set of specific tautologies. In particular, Lutz Strasburger introduced in [2] the notion of balanced formulas and showed that if there are polynomially bounded Frege proofs for the set of balanced tautologies, then the Frege systems are super. An analogous result for some other class of tautologies is proved in [3].

In this work, we introduce formulas that can be described by sign-alternating trees (**sat** formulas) and show that the proofs of tautologies that contain only $\supset$ and $\neg$ symbols, where $\neg$ is used only in literals, can be polynomially reduced to proofs of specific formulas constructed from **sat** formulas.

2. Main Notions and Notations

We will use the current concepts of a classical tautology, Frege proof systems for classical propositional logic, proof and proof complexity [1]. Let us recall some of them.

A Frege system $\mathcal{F}$ uses a denumerable set of propositional variables, a finite, complete set of propositional connectives; $\mathcal{F}$ has a finite set of inference rules defined by a figure of the form $\frac{A_1 A_2 ... A_n}{B}$ (the rules of inference with zero hypotheses are the axioms schemes); $\mathcal{F}$ must be sound and complete, i.e., for each rule of inference $\frac{A_1 A_2 ... A_n}{B}$ every truth-value assignment, satisfying $A_1 A_2 ... A_n$, also satisfies B, and $\mathcal{F}$ must prove every tautology.

The particular choice of a language for the presented propositional formulas is immaterial in this consideration. However, for some technical reasons, we assume that the language contains propositional variables, logical connectives $\neg, \wedge, \vee, \supset$ and parentheses $(,)$. Note that some parentheses can be omitted in generally accepted cases.

By $|\varphi|$ we denote the size of a formula φ, defined as the number of entries of all logical signs in it. It is obvious that the full size of a formula, which is understood to be the number of all symbols is bounded by some linear function in $|\varphi|$.

In the theory of proof complexity, the two main characteristics of the proof are: t-complexity (length), defined as the number of proof steps, l-complexity (size), defined as the sum of sizes for all formulas in the proof (formal definitions are, for example, in [4]).

Let ϕ be a proof system and φ be a tautology. We denote by $l^{\phi}_{\varphi}(t^{\phi}_{\varphi})$ the minimal possible value of l-complexity (t-complexity) for all ϕ-proofs of tautology φ.

Let M be some set of tautologies.

Definition 1: *We call the ϕ-proofs of tautologies from a set M l-polynomially (t-polynomially) bounded if there is a polynomial p such that $l^{\phi}_{\varphi} \leq p(|\varphi|)$ ($t^{\phi}_{\varphi} \leq p(|\varphi|)$) for all φ from M.*

Definition 2: *We call the ϕ-proofs of tautologies from a set M l-linearly (t-linearly) bounded if there is a linear function f such that $l^{\phi}_{\varphi} \leq f(|\varphi|)$ ($t^{\phi}_{\varphi} \leq f(|\varphi|)$) for all φ from M.*

Now we'll give the definition of **sat** formulas and prove some lemmas, which are necessary for proving the main result.

Definition 3: *We'll say that a formula is described by a sign-alternating tree (**sat** formula) if it satisfies the following rules.*

1. *it's a literal*
2. *has a form $r \wedge (T_1 \vee T_2)$, where r is a literal and T_1, T_2 are **sat** formulas*

Lemma 1: *For any formulas A, B, C, the following formulas have polynomially bounded proofs.*

1. $A \equiv (C \supset A) \wedge (\neg C \supset A)$
2. $A \supset (B \supset A)$
3. $(\neg A \supset (B \supset A)) \equiv (\neg A \supset \neg B)$
4. $\neg\neg A \equiv A$

5. $A \supset (B \supset C) \equiv A \wedge B \supset C$
6. $A \wedge \neg A \wedge B \supset C$
7. $A \wedge B \supset A$
8. $\neg(A \supset B) \equiv A \wedge \neg B$
9. $A \supset (B \wedge C) \equiv (A \supset B) \wedge (A \supset C)$
10. $A \supset (B \supset C) \equiv B \supset (A \supset C)$
11. $A \supset (A \supset B) \equiv A \supset B$
12. $(A \supset B) \wedge (C \supset B) \equiv (A \vee C) \supset B$

The proof is trivial as all the formulas are tautologies and have fixed length proofs, so the proof complexities may be assumed to be linearly bounded.

Lemma 2: *Tautologies of the form*

$$A \supset (B_1 \supset ...(B_{n-1} \supset (B_n \supset A))...)$$

have polynomially bounded proofs.

Proof. We can prove the tautology above by the following steps.

$$\begin{array}{ll}
A \vdash A & \\
A \vdash A \supset (B_n \supset A) & \text{2nd formula of Lemma 1} \\
A \vdash (B_n \supset A) & \textit{modus ponens} \\
A \vdash (B_n \supset A) \supset (B_{n-1} \supset (B_n \supset A)) & \\
A \vdash (B_{n-1} \supset (B_n \supset A)) & \textit{modus ponens} \\
\vdots & \\
A \vdash (B_1 \supset ...(B_{n-1} \supset (B_n \supset A))...) &
\end{array}$$

The number of proof steps is linearly bounded and the size of each formula in proof is also linearly bounded, so the proof is polynomially bounded. ■

Lemma 3: *Tautologies of the form*

$$d_1 \supset (d_2 \supset (... \supset d_k)...)$$

where $d_1, d_2, ..., d_k$ are literals, have polynomially bounded proofs.

Proof. After applying the operation of replacement by an equivalent formula $(k-2)$ times using the 5th formula of Lemma 1, we get:

$$d_1 \supset (d_2 \supset (... \supset d_k)...) \equiv d_1 \wedge d_2 \wedge ... \wedge d_{k-1} \supset d_k$$

In this case, if there exist $1 \leq i, j, \leq k-1$ such that $d_i = \neg d_j$ then replacing them with equivalent formulas using the 6th formula of Lemma 1, we'll get a polynomially bounded proof, or if there exist such $1 \leq i \leq k-1$ such that $d_i = d_k$ then replacing it with equivalent formulas using the 7th formula of Lemma 1, we'll get a polynomially bounded proof, otherwise the formula isn't a tautology. ■

3. Main Result

Definition 4: *Any propositional formula A is called sat-constructed if it is in the following form: $A = \neg(T_1 \wedge T_2 \wedge ... \wedge T_n)$, where $T_i(1 \leq i \leq n)$ are **sat** formulas.*

Theorem 1: *Let M be the set of all sat-constructed tautologies. If proofs of formulas from the set M are l-polynomially (t-polynomially) bounded, then proofs of all tautologies containing only $\supset$ and $\neg$ symbols, where $\neg$ is used only in literals, are l-polynomially (t-polynomially) bounded.*

Proof. We need to prove that any tautology A containing only $\supset$ and $\neg$ symbols, where $\neg$ is used only in literals, can be reduced to a *sat-constructed* tautology in polynomially bounded number of steps and length.

If the number of implications in the formula is bounded by, let's say, 3, then it can be reduced to $\neg(p_1 \wedge \neg p_1)$, and the reduction complexity will be constant which is also a polynomial. We now assume that formula A contains more than 3 implications. A can be expressed in the following form:

$$A = (S_1 \supset ...(S_{c-1} \supset (S_c \supset q_1))...) \tag{1}$$

where $S_i(1 \leq i \leq c)$ are sub-formulas and q_1 is a literal. We can replace A with an equivalent formula using the 1st formula of Lemma 1 and get $(q_1 \supset A) \wedge (\neg q_1 \supset A)$. The first half has a polynomially bounded proof by Lemma 2.

For the second half, we can apply the operation of replacement by an equivalent formula using the 3rd formula of Lemma 1 and get:

$$\neg q_1 \supset (S_1 \supset ...(S_{c-1} \supset (S_c \supset q_1))...) \equiv \neg q_1 \supset (S_1 \supset ...(S_{c-1} \supset \neg S_c)...)$$

So the proof of A has been polynomially reduced to the proof of $\neg q_1 \supset (S_1 \supset ...(S_{c-1} \supset \neg S_c)...)$. If S_c is a literal, then we can repeat the same process for the formula we got. After some repetitions, we'll end up either with a tautology $d_1 \supset (d_2 \supset (... \supset d_k)...)$, where $d_i(1 \leq i \leq k)$ are literals. This tautology has a polynomially bounded proof by Lemma 3. Or we'll end up with the following formula:

$$d_1 \supset (d_2 \supset ... \supset (d_m \supset (U_1 \supset ...(U_k \supset \neg F)...))...) \tag{2}$$

where $d_i(1 \leq i \leq m)$ are literals and F is not a literal.

Suppose $F = F_1 \supset F_2$, applying the replacement by an equivalent formula using the 8th formula of Lemma 1, we'll get:

$$\begin{gathered} d_1 \supset (d_2 \supset ... \supset (d_m \supset (U_1 \supset ...(U_k \supset \neg F)...))...) \equiv \\ d_1 \supset (d_2 \supset ... \supset (d_m \supset (U_1 \supset ...(U_k \supset (F_1 \wedge \neg F_2))...))...) \end{gathered}$$

Applying the operation of replacement by an equivalent formula multiple times using the 9th formula of Lemma 1, we'll get:

$$\begin{gathered} d_1 \supset (d_2 \supset ... \supset (d_m \supset (U_1 \supset ...(U_k \supset (F_1 \wedge \neg F_2))...))...) \equiv \\ d_1 \supset (d_2 \supset ... \supset (d_m \supset (U_1 \supset ...(U_k \supset F_1)...))...) \wedge \\ d_1 \supset (d_2 \supset ... \supset (d_m \supset (U_1 \supset ...(U_k \supset \neg F_2)...))...) \end{gathered}$$

So, after polynomially bounded number of steps, we reduced the proof of A to the proof of two formulas, the first one of which has the form (1) and the second one has the form (2). Repeating the same process for these formulas, we'll get l tautologies, and the proof of A will be reduced to the proof of these tautologies:

$$\begin{array}{c} d_1^1 \supset (d_2^1 \supset ... \supset (d_{m_1}^1 \supset (d_1 \supset (d_2 \supset ... \supset (d_m \supset (U_1 \supset ... \supset \neg U_k))...)))...) \\ d_1^2 \supset (d_2^2 \supset ... \supset (d_{m_2}^2 \supset (d_1 \supset (d_2 \supset ... \supset (d_m \supset (U_1 \supset ... \supset \neg U_k))...)))...) \\ \vdots \\ d_1^l \supset (d_2^l \supset ... \supset (d_{m_l}^l \supset (d_1 \supset (d_2 \supset ... \supset (d_m \supset (U_1 \supset ... \supset \neg U_k))...)))...) \end{array}$$

where $d_1^i, d_2^i, ..., d_{m_i}^i (1 \leq i \leq l)$ are literals. There are no repetitions among $d_1^i, d_2^i, ..., d_{m_i}^i, d_1, ..., d_m$, otherwise we can keep a single one of each repetition - replacing by equivalent formulas using the 10th and 11th formulas of Lemma 1. We may also assume that there are no variable repetitions, because if a variable and it's negation are present, then the formula can be polynomially proved. So we can assume that all the literals among $d_1^i, d_2^i, ..., d_{m_i}^i, d_1, ..., d_m$ are different and all the variables are also different and so the number of those literals doesn't exceed $|A|$. Also note that each time a new formula was generated an implication from F was removed, so $l < |F| < |A|$.

Applying the operation of replacement by an equivalent formula using the 10th formula of Lemma 1, we'll get the following form for our l formulas:

$$\begin{array}{c} C_1 = d_2 \supset (d_3 \supset ... \supset (d_m \supset (d_1 \supset (d_1^1 \supset ... \supset (d_{m_1}^1 \supset (U_1 \supset ... \supset \neg U_k))...)))...) \\ C_2 = d_2 \supset (d_3 \supset ... \supset (d_m \supset (d_1 \supset (d_1^2 \supset ... \supset (d_{m_2}^2 \supset (U_1 \supset ... \supset \neg U_k))...)))...) \\ \vdots \\ C_l = d_2 \supset (d_3 \supset ... \supset (d_m \supset (d_1 \supset (d_1^l \supset ... \supset (d_{m_l}^l \supset (U_1 \supset ... \supset \neg U_k))...)))...) \end{array}$$

At this point, we've reduced the proof of A to the proof of $C_1 \wedge C_2 \wedge ... \wedge C_l$ polynomially.

Applying the operation of replacement by an equivalent formula using the 9th formula of Lemma 1, we'll get:

$$C_1 \wedge C_2 \wedge ... \wedge C_l \equiv d_2 \supset (d_3 \supset ... \supset (d_m \supset (C_1' \wedge C_2' \wedge ... \wedge C_l'))...)$$

where

$$\begin{array}{c} C_1' = (d_1 \supset (d_1^1 \supset ... \supset (d_{m_1}^1 \supset (U_1 \supset ... \supset \neg U_k))...)) \\ C_2' = (d_1 \supset (d_1^2 \supset ... \supset (d_{m_2}^2 \supset (U_1 \supset ... \supset \neg U_k))...)) \\ \vdots \\ C_l' = (d_1 \supset (d_1^l \supset ... \supset (d_{m_l}^l \supset (U_1 \supset ... \supset \neg U_k))...)) \end{array}$$

Applying the operation of replacement by an equivalent formula using the 5th formula of Lemma 1, we'll get the following form for above formulas:

$$\begin{array}{c} C_1' = d_1 \wedge d_1^1 \wedge ... \wedge d_{m_1}^1 \supset (U_1 \supset ... \supset \neg U_k) \\ C_2' = d_1 \wedge d_1^2 \wedge ... \wedge d_{m_2}^2 \supset (U_1 \supset ... \supset \neg U_k) \\ \vdots \\ C_l' = d_1 \wedge d_1^l \wedge ... \wedge d_{m_l}^l \supset (U_1 \supset ... \supset \neg U_k) \end{array}$$

Applying the operation of replacement by an equivalent formula using the 12th formula of Lemma 1, we'll get:

$$C'_1 \wedge C'_2 \wedge ... \wedge C'_l \equiv (d_1 \wedge d_1^1 \wedge ... \wedge d_{m_1}^1 \vee d_1 \wedge d_1^2 \wedge ... \wedge d_{m_1}^2 \vee ... \vee d_1 \wedge d_1^l \wedge ... \wedge d_{m_1}^l) \supset (U_1 \supset ... \supset \neg U_k)$$

Let's prove that the **DNF** generated above - $(d_1 \wedge d_1^1 \wedge ... \wedge d_{m_1}^1 \vee d_1 \wedge d_1^2 \wedge ... \wedge d_{m_1}^2 \vee ... \vee d_1 \wedge d_1^l \wedge ... \wedge d_{m_1}^l)$ is an **sat** formula.

As we can see, each of the conjunctions includes d_1. Note that we can clearly split the **DNF** into two **subDNF**'s - one generated from F_1 and the other generated from F_2. For each of F_1 and F_2, we repeated the same process and so if the generated **DNF**s from F_1 and F_2 are both **sat** formulas, then the **DNF** generated above is also an **sat** formula. Note that we can make the assumption above, as after finite repetitions of splitting operation, we'll reach a literal, which is an **sat** formula.

At this point, we have polynomially reduced the proof of tautology A to the proof of the following tautology:

$$d_2 \supset (d_3 \supset ... \supset (d_m \supset (d_1 \wedge d_1^1 \wedge ... \wedge d_{m_1}^1 \vee d_1 \wedge d_1^2 \wedge ... \wedge d_{m_1}^2 \vee ... \vee d_1 \wedge d_1^l \wedge ... \wedge d_{m_1}^l) \supset (U_1 \supset ... \supset \neg U_k)...)$$

By deduction theorem the proof of the above formula is equivalent to the following proof:

$$d_2, ..., d_m, (d_1 \wedge d_1^1 \wedge ... \wedge d_{m_1}^1 \vee d_1 \wedge d_1^2 \wedge ... \wedge d_{m_1}^2 \vee ... \vee d_1 \wedge d_1^l \wedge ... \wedge d_{m_1}^l) \vdash (U_1 \supset ... \supset \neg U_k)...)$$

Note that all the hypotheses are **sat** formulas and that we can repeat all the previous steps on formula $(U_1 \supset ... \supset \neg U_k)$ even though we have some hypotheses. Repeating this process, the proof of A will be reduced to the proof of the following formula:

$$T_1 \wedge T_2 \wedge ... \wedge T_n \supset c_1,$$

where $T_i (1 \leq i \leq n)$ are **sat** formulas and c_1 is a literal. Note that new T_i-s are generated only when we consider the last sub-formula of A and then remove it for the next step, so this guarantees that $n \leq |A|$.

Applying the operation of replacement by an equivalent formula using the 1st formula of Lemma 1, we'll get:

$$T_1 \wedge T_2 \wedge ... \wedge T_n \supset c_1 \equiv (c_1 \supset (T_1 \wedge T_2 \wedge ... \wedge T_n \supset c_1)) \wedge (\neg c_1 \supset (T_1 \wedge T_2 \wedge ... \wedge T_n \supset c_1))$$

The tautology $(c_1 \supset (T_1 \wedge T_2 \wedge ... \wedge T_n \supset c_1))$ has a polynomially bounded proof.

We reduced the proof of A to the proof of $(\neg c_1 \supset (T_1 \wedge T_2 \wedge ... \wedge T_n \supset c_1))$. Applying the operation of replacement by an equivalent formula using the 3rd formula of Lemma 1, we'll get:

$$A \equiv \neg(T_1 \wedge T_2 \wedge ... \wedge T_n)$$

All the operations have polynomial complexity. The number of steps is also polynomially bounded, so the total complexity of reduction is polynomially bounded. ■

4. Conclusion

In this work, we introduced **sat** formulas and reduced the proofs of tautologies containing only $\supset$ and $\neg$ symbols, where $\neg$ is used only in literals, to the proofs of *sat-constructed* tautologies in polynomially bounded number of steps and length. Investigation of proof complexities of *sat-constructed* tautologies is in process.

References

[1] S. A. Cook and A. R. Reckhow, "The relative efficiency of propositional proof systems", *Journal of Symbolic logic*, vol. 44, pp. 36-50, 1979.

[2] L. Strasburger, "Extension without Cut", *Annals of Pure and Applied Logic*, vol. 163, no. 12, pp. 1995-2007, 2012.

[3] A. A. Chubaryan and G. V. Petrosyan, "Some notes on proof complexities in Frege systems", *Sciences of Europe*, vol 1. # 12 (12), Physics and Mathematics, pp. 31–34, 2017.

[4] J. Nordstrom, "Narrow proofs may be spacious: Separating space and width in resolution", *SIAM Journal on Computing*, vol. 39, no. 1, pp. 59-121, 2019.

Որոշ տիպի նույնաբանությունների արտածման բարդությունների վերաբերյալ

Վահագն Ն. Ալթունյան և Գարիկ Վ. Պետրոսյան

Երևանի պետական համալսարան
e-mail: altunyanv@gmail.com, garik.petrosyan.1@gmail.com

Ամփոփում

Այս հոդվածում ուսումնասիրվում են արտածման բարդությունները հատուկ տիպի նույնաբանությունների համար, որոնք նկարագրվում եմ որպես իմպլիկացիաներով և լիտերալներով կազմված նույնաբանություններ: Մասնավորապես ապացուցվել է, որ այդ տեսքի նույնաբանությունների արտածումները բազմանդամորեն հանգեցվում են նշանափոխ ծառերով ներկայացվող նույնաբանություն հանդիսացող բանաձևերի արտածումներին:

Բանալի բառեր՝ Ֆրեգեի համակարգեր, նույնաբանություններ, նշանափոխ ծառեր, արտածման բարդություն:

О сложности выводов некоторого типа тавтологий

Ваагн Н. Алтунян и Гарик В. Петросян

Ереванский государственный университет
e-mail: altunyanv@gmail.com, garik.petrosyan.1@gmail.com

Аннотация

В настоящей статье исследованы сложности выводов тавтологий специального вида, которые можно описать как тавтологии состоящие из импликаций и литералов. В частности, доказано, что выводы тавтологий такого вида можно полиномиально свести к выводам тавтологий, которые являются формулами, описываемыми знакопеременными деревьями.

Ключевые слова: системы Фреге, тавтологии, знакопеременные деревья, сложность вывода.

Կանոններ հեղինակների համար

ՀՀ ԳԱԱ ԻԱՊԻ “Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ” պարբերականը տպագրվում է 1963 թվականից: Պարբերականում հրատարակվում են նշված ոլորտին առնչվող գիտական հոդվածներ, որոնք պարունակում են նոր՝ չհրատարակված արդյունքներ:

Հոդվածները ներկայացվում են անգլերեն՝ ձևավորված համապատասխան “ոճով” (style): Հոդվածի ձևավորման պահանջներին ավելի մանրամասն կարելի է ծանոթանալ պարբերականի կայքէջում՝ http://mpcs.sci.am/:

Rules for authors

The periodical “Mathematical Problems of Computer Science” of IIAP NAS RA has been published since 1963. Scientific articles related to the noted fields with novel and previously unpublished results are published in the periodical.

Papers should be submitted in English and prepared in the appropriate style. For more information, please visit the periodical's website at http://mpcs.sci.am/.

Правила для авторов

Журнал «Математические проблемы компьютерных наук» ИПИА НАН РА издается с 1963 года. В журнале публикуются научные статьи в указанной области, содержащие новые и ранее не опубликованные результаты.

Статьи представляются на английском языке и оформляются в соответствующем стиле. Дополнительную информацию можно получить на веб-сайте журнала: http://mpcs.sci.am/.

The electronic version of the periodical “Mathematical Problems of Computer Science” and rules for authors are available at

http://mpcs.sci.am/

Phone: (+37460) 62-35-51
Fax: (+37410) 28-20-50
E-mail: mpcs@sci.am
Website: http://mpcs.sci.am/

Ստորագրված է տպագրության՝ 14.12.2021
Թուղթը՝ օֆսեթ:
ՀՀ ԳԱԱ «Գիտություն» հրատարակչություն
Ծավալը՝ 74 էջ: Տպաքանակը՝ 100
ՀՀ ԳԱԱ ԻԱՊԻ Համակարգչային պոլիգրաֆիայի լաբորատորիա
Երևան, Պ. Սևակի 1
Հեռ. +(374 60) 623553
Գինը՝ անվճար

Подписано в печать 14.12.2021
Офсетная бумага.
Издательство «Гитутюн» НАН РА
Объём: 74 страниц. Тираж: 100
Лаборатория компьютерной
полиграфии ИПИА НАН РА.
Ереван, П. Севака 1
Тел.: +(374 60) 623553
Цена: бесплатно

Signed in print 14.12.2021
Offset paper
“Gitutyun” publishing house NAS RA
Volume: 74 pages
Circulation: 100
Computer Printing Lab
of IIAP NAS RA
Yerevan, 1, P. Sevak str.
Phone: +(374 60) 623553
Free of charge