RESEARCH AND DEVELOPMENT OF DEEP LEARNING MODELS FOR THE RECOGNITION OF ARMENIAN HANDWRITING LETTERS *

UDC 004.89:81'351(479.25)

Karen Nikoghosyan
2nd year master student of the "System Programming" department of the Armenian-Russian University, programmer of "Scylla" company
karen.nikoghosyan.98@gmail.com
Yerevan, Armenia

Qajik Hakobyan
YSU Faculty of Applied Mathematics, 2nd-year master student of "Digital Analysis" department, engineer of "Krisp" company
qajikhakobyan@yahoo.com
Yerevan, Armenia

Nowadays, handwritten letter recognition(HLR) is a very interesting and challenging task. This task remains challenging because, there are many problems connected with handwriting, for example, every people have a different handwriting style, and it can be changed over time. In contrast to printed text, which sits straight, handwritten text can rotate to the right, or it can be broken, and so on. There are many approaches for solving that task, but the most popular and effective method is deep learning. The purpose of this article is to find the best Artificial Intelligence (AI) based solution for Armenian handwriting letters recognition. As the Armenian language is not so popular, and its usage is limited, there aren't any AI-based good models or tools that will recognize Armenian handwriting.

This work uses the most popular, effective, and modern methods of deep learning. The Mashtots database, recently created in the Basic Research Laboratory of "Automation Systems & Modeling" of the National Polytechnic University of Armenia, was used.

Keywords: *Mashtots dataset, Visual pattern recognition, Convolutional Neural Network (CNN), Spinal Network, Artificial Intelligence (AI).*

Economic significance

In simple terms, HLR (Handwritten Letter Recognition) is the process of converting text in images into text format. The main application of HLR technology is found in various tasks related to data digitization. It can extract information from documents, invoices, ID cards.

Nowadays, Artificial Intelligence-based applications and products are widely integrated into almost any company. They help people to save their time and company owner for saving money. Being a part of Artificial Inleniganese problems, HLR can be integrated into products to automate many jobs.

---

* Հոդվածը ներկայացվել է 22.03.2022:
Տպագրության է երաշխավորել ՀՊՃՀ դասախոս, «Սինոփսիս» ընկերության ավագ ծրագրավորող, տ.գ.թ. Ա. Պետրոսյանը: 29.02.2022:
Հոդվածը տպագրության է ընդունվել 29.03.2022:

Various advantages of handwritten letter recognition technology have helped businesses save time by manually inserting data into a device, improving job management, reducing the cost of translating documents into digital form, and, among others, minimizing manual errors. In addition, it offers other benefits such as improved customer support and increased document protection.

Introduction

The handwriting letter recognition(HLR)[1] process allows the conversion of various types of documents into parsable, editable, and searchable data. The ultimate goal of HLR is to make a machine capable of reading, editing, interacting with the content just like a human in a short amount of time.

As a result of many practical applications in day-to-day activities, handwritten letter recognition has become increasingly important in today's digitized world. Several recognition systems have been developed in recent years to use in areas where high classification efficiencies are required. To solve more complex tasks that would otherwise be tedious and expensive, handwriting recognition systems are used to recognize letters, characters, and digits in handwriting. Banks use automated processing systems to process bank checks. Without automated computers, the bank would require a lot of employees who would not be as efficient as computerized systems. Biological neural networks can inspire handwriting recognition systems, which allow humans and animals to learn and model complex relationships. The artificial neural network[8] can be used to generate them.

People can recognize digits, letters, and characters in handwriting through their brains. Humans are biased so they may turn to different interpretations of handwriting. In contrast, they are unbiased and can accomplish very challenging tasks that humans may spend a lot of time and energy on if similar tasks have to be done by humans. Understanding how human-readable underwriting works is important. The human visual system is primarily involved when reading handwriting characters, letters, words, or digits. Because it appears effortless, reading handwriting is not as easy as it seems. Although everything happens unconsciously, humans can make sense of what they see based on what their brains have been taught. Handwriting may not seem difficult to humans.

Developing a computer system to read handwriting reveals the challenge of visual pattern recognition[2]. Handwriting recognition systems are best developed using artificial neural networks. They simulate how the human brain works when reading handwriting with neural networks that simplify the process. Machines are capable of matching and even surpassing human abilities in this area. People have different handwriting styles, some of which are difficult to read. In addition, reading handwriting can be time-consuming and tedious, especially when people have to read multiple handwritten documents of different persons.

One of the deep learning architectures that can be used to recognize an object in a digital image is the Convolutional Neural Network (CNN) [4]. CNN is a type of Neural Network specifically for processing data with a grid-shaped topology. CNN is the best model in handling object detection and object recognition.

Handwritten recognition performance has improved dramatically in recent years, but so far, handwritten recognition remains a challenging task. There are many problems with handwriting.

- *A person's handwriting style varies from time to time and is inconsistent.*
- *A document or image that has degraded over time.*

- *In contrast to printed text, which sits straight, handwritten text can rotate to the right.*
- *It is expensive to collect a well-labeled dataset to learn compared to synthetic data.*
- *Different characteristics of each person's handwriting*
- *Some characters have similar shapes*
- *Broken or distorted characters*
- *Differences in the thickness of the characters written*

The Armenian language is one of the hardest languages, they have 38 letters.

The purpose of this article is to find the best modern and effective AI model architecture for Armenian handwritten letters. For that, we use the most popular ResNet[7], VGG[6] architectures, and also use SpinalNet[3]. The use of SpnalNet architecture is due to two reasons, first, it is released recently, and second SpinalNet classification layers provided the state-of-the-art (SOTA) performance on QMNIST, EMNIST (Letters, Digits, and Balanced) datasets. The results can be found at the link below. https://paperswithcode.com/sota/image-classification-on-emnist-letters

Material and methods

Dataset: *The dataset is considered to be one of the important parts of any artificial intelligence problem. The accuracy of the model can change significantly depending on the quality of the dataset and the amount of data in it.*

In this article, we use the Mashtots dataset[5] which was created in the Basic Research Laboratory of "Automation Systems & Modeling" of the National Polytechnic University of Armenia. Further additions and corrections were made at the Science and Technology Foundation of Armenia (FAST).

The training set of "Mashtots" data set consists of 70060 pictures (64x64 size), which are distributed in 78 folders. The set of the test is presented in new_test.csv and contains 50,000 data, most of which are false data to exclude fraud (manual labeling). Figure 1 shows some letters example from the Mashtots dataset.

Figure 1. Examples from Mashtots dataset

Data augmentation and Traning: *All AI models for achieving good results need to see a huge amount of data. But when we have a limited number of items in the dataset we need to apply some data augmentation techniques to get good performance. Furthermore, the number of parameters your model requires is proportional to the complexity of the task it must perform. The* *example of how data augmentation work shows in Figure 2. We apply two techniques to our dataset: RandomPerspective and RandomRotation from the PyTorch python library.*

For all models use train 200 epoch, used Adam optimizer[6], and also cross-entropy loss function[7]. Initially, we set the learning rate at 0.005.

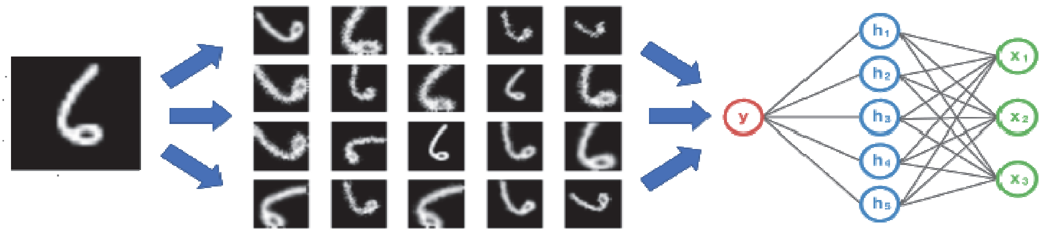

Figure 2. Example of data augmentation

https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/

***VGG:** VGG16 is a convolution neural net (CNN) architecture. One of the unique things about VGG16 is that it is based on convolution layers of a 3x3 filter with a stride 1, and it uses the same padding and maximum pool layer of a 2x2 filter with a stride 2. In the whole architecture, the convolution and maximum pool layers are arranged consistently. A final output layer consists of two FCs (fully connected layers) and a softmax. There are 16 layers in VGG16, each with its own weight. It has 138 million (approximately) parameters and is quite a large network. The architecture of VGG16 is shown below.*

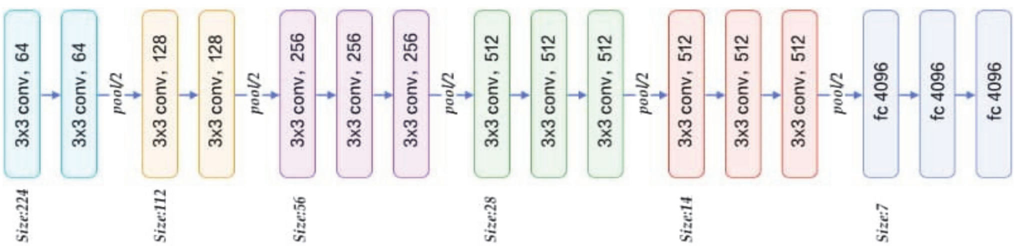

Figure 3: The VGG16 architecture https://iq.opengenus.org/vgg16/

***VGGSpinal:** After the first training with VGG16, we apply Spinal layers. Spinal layers were applied instead of the last fully connected layers. The network architecture of VGGSpinal shows below, in Figure 4.*

Figure 3: The VGGSpinal architecture

VGGSmall: *We take into count the factor that VGG16 has a pretty large network, and as we traying to solve not so hard problem, we made another model by cutting some layers from VGG16 and calling it VGGSmall. We cut the last convolutional layer, and also change input counts to 256. The architecture of VGGSmall shows below.*

Figure 3: The VGGSmall architecture

VGGSmallSpinal: *The test results of VGGSmall show that it works much better than VGG16, and we decide to add Spinal layers in VGGSmall instead of VGG16. We built the model by replacing the last VGGSmalls dense layers with the Spinal layer. The network architecture of the VGGSmallSpinal model shows below.*

Figure 3: The VGGSmallSpinal architecture

ResNet: *ResNet or residual networks are artificial neural networks that help to build a deeper neural network by utilizing shortcuts or skip connections in order to jump over layers.*

ResNet has different versions, such as ResNet-18, ResNet-34, ResNet-50, and so on. The numbers refer to layers, even though the architecture is the same. To put it simply, ResNet's main component is the residual block. The block structure is shown below:

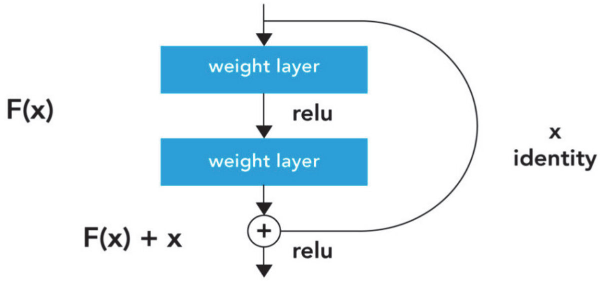

Figure 2. Residual building block

We can see that there is a direct connection that skips some layers (may vary in different models) in between. This connection is known as the 'skip connection' and forms the basis of residual blocks. Because of this skip connection, the output of the layer is different now. In the absence of this skip connection, the input 'x' is multiplied by the weights of the layer, followed by a bias term. Next, we pass this term through the activation function, f(), and we get our output as H(x).

The mathematical equation of identity mapping with the residual network is given below:

$$y = F(x,\{Wi\}) + x$$

Formula 1. The mathematical equation of identity mapping

In this work, we use RestNet18 architecture, and also the ResNet with SpinalNet.

The architecture of ResNet18 is shown in Figure 5. We don't go deeper into how ResNet works. You can find more about ResNet in this article[].

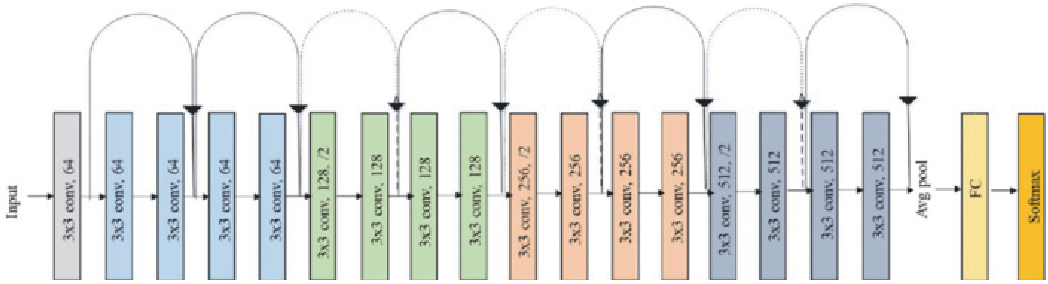

Figure 5: The ResNet18 architecture

https://www.researchgate.net/figure/Original-ResNet-18-Architecture_fig1_336642248

ResNetSpinal: *For the ResNetSpinal model we replace the last fully connected layer with the Spinal layer. The network architecture of the ResNetSpinal model shows in Figure 6*

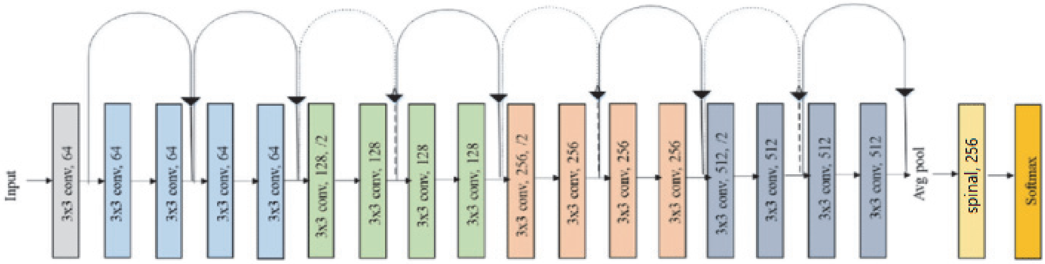

Figure 6: The ResNetSpinal architecture

Results and analysis

All models were trained on Nvidia 3060 RTX GPU. During the first 10 epochs of training, results showed the accuracy of our models was not getting better, and we reduced the learning rate 2 times.

The training shows that the best result gives the VGGSmallSpinal model. The resulting graphic is given in Figure 7. This result was predictable because handwritten recognition is not required for very deep neural networks, in that case, VGGSmall is the best choice. Also applying Spinal nets on the VGGSmall network as the fully connected layer increases the classification result. The research resulted in finding the most efficient AI model to recognize Armenian handwriting.

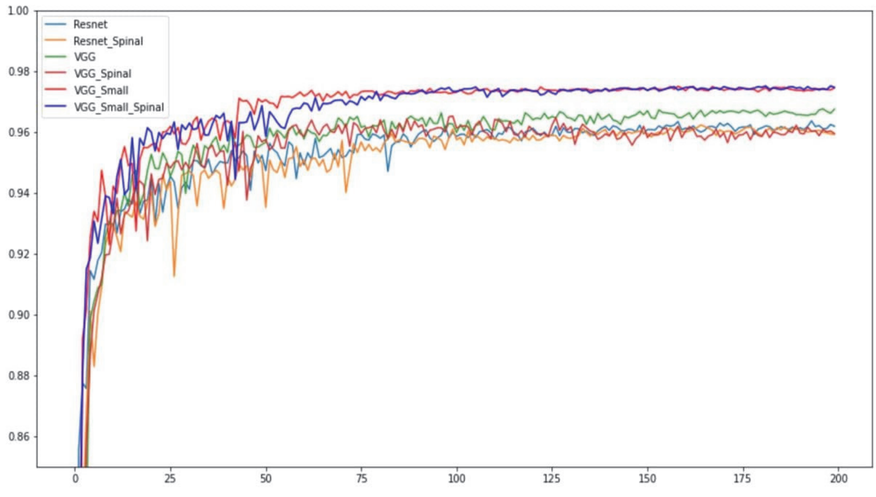

Figure 7. Graphical representation of the accuracy of all models

The results are shown in Table 1 presents the accuracy of VGGSmallSpinal being equal to 0.97515 which is yet the best among the previous models.

Model	**Accuracy**
VGG	0.96762
VGGSmall	0.97503
VGGSpinal	0.96523
VGGSmallSpinal	**0.97515**
ResNet	0.96368
RestNetSpinal	0.96236

Table 1. Accuracy of all models

Conclusion

During the research, 6 types of deep learning medals was created. The models are based on existing known, efficient, and new artificial intelligence algorithms.

After testing the models, it became clear that the VGGSmallSpinal model showed the best results. The model was based on the very popular VGG neural network and Spinal Networks. The result is a new type of model, which has surpassed other models in terms of efficiency.

REFERENCES

1. Eden, M. (1962). Handwriting and pattern recognition. IRE Transactions on Information Theory, 8(2), 160-166.
2. Fukushima, K. (1988). A neural network for visual pattern recognition. Computer, 21(3), 65-75.
3. Kabir, H. M., Abdar, M., Jalali, S. M. J., Khosravi, A., Atiya, A. F., Nahavandi, S., & Srinivasan, D. (2020). Spinalnet: Deep neural network with gradual input. arXiv preprint arXiv:2007.03347.
4. Kim, P. (2017). Convolutional neural network. In MATLAB deep learning (pp. 121-147). Apress, Berkeley, CA.
5. Mashtoc dataset. Created in the Basic Research Laboratory of "Automation Systems & Modeling" of the National Polytechnic University of Armenia
6. Mehta, S., Paunwala, C., & Vaidya, B. (2019, May). CNN based traffic sign classification using adam optimizer. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)* (pp. 1293-1298). IEEE.

7. Li, L., Doroslovački, M., & Loew, M. H. (2020). Approximating the gradient of cross-entropy loss function. *IEEE Access*, *8*, 111626-111635.
8. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
9. Targ, S., Almeida, D., & Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. arXiv preprint arXiv:1603.08029.
10. Wang, S. C. (2003). Artificial neural network. In Interdisciplinary computing in java programming (pp. 81-100). Springer, Boston, MA.

РЕЗЮМЕ

Исследование и разработка моделей глубокого обучения для распознавания армянских букв рукописного ввода

Карен Никогосян

магистрант 2 курса кафедры "Системное программирование" Армяно-Российского Университета, программист компании "Scylla",

Ереван, Армения

Каджик Акопян

Факультет прикладной математики ЕГУ, магистрант 2 курса кафедры "Цифровой анализ", инженер компании "Крисп"

Ереван, Армения

Ключевые слова: Набор данных Mashtots, Распознавание визуальных образов, Сверточный нейронная сеть, Спинальная сеть, Искусственный интеллект (ИИ).

В настоящее время распознавание рукописных букв является очень интересной и сложной задачей. Эта задача остается сложной, потому что с почерком связано много проблем, например, почерк у всех людей разный, и со временем он может измениться. В отличие от печатного текста, который расположен прямо, рукописный текст может поворачиваться вправо, ломаться и т. д. Существует множество подходов к решению этой задачи, но наиболее популярным и эффективным методом является глубокое обучение. Цель этой статьи — найти лучшее решение на основе искусственного интеллекта (ИИ) для распознавания армянских рукописных букв. Поскольку армянский язык не так популярен и его использование ограничено, не существует хороших моделей или инструментов на основе ИИ, которые распознают армянский почерк.

В данной работе используются самые популярные, эффективные и современные методы глубокого обучения. Использовалась база данных Маштоца, недавно созданная в Исследовательской лаборатории «Системы автоматизации и моделирование» Национального политехнического университета Армении.

ԱՄՓՈՓՈՒՄ

Հայերեն ձեռագիր տառերի ճանաչման խորը ուսուցման մոդելների հետազոտություն և մշակում

Կարեն Նիկողոսյան

Հայ-ռուսական համալսարանի «Համակարգային Ծրագրավորում» բաժնի
մագիստրատուրայի 2-րդ կուրսի ուսանող,
«Սցիլլա» ընկերության ծրագրավորող
Երևան, Հայաստան

Քաջիկ Հակոբյան

ԵՊՀ Կիրառական Մաթեմատիկա Ֆակուլտետի
«Թվային Անալիզ» բաժնի մագիստրատուրայի
2-րդ կուրսի ուսանող, «Կրիսպ» ընկերության ինժեներ
Երևան, Հայաստան

Բանալի բառեր՝ Mashtots տվյալների բազա, Տեսողական օրինաչափությունների ճանաչում, Փաթույթային նեյրոնային ցանց, Ողնաշարային ցանցեր, Արհեստական բանականություն:

Մեր օրերում ձեռագիր տառերի ճանաչումը շատ հետաքրքիր և բարդ խնդիր է: Այս խնդիրը մնում է դժվար, քանի որ ձեռագրի հետ կապված բազմաթիվ խնդիրներ կան, օրինակ՝ յուրաքանչյուր անձ ունի ձեռագրի տարբեր ոճ, և այն կարող է փոխվել ժամանակի ընթացքում: Ի տարբերություն տպագիր տեքստի, որը ուղիղ է, ձեռագիր տեքստը կարող է պտտվել դեպի աջ, կամ կարող է կտրտված լինել և այլն: Այդ խնդիրը լուծելու բազմաթիվ մոտեցումներ կան, սակայն ամենահայտնի և արդյունավետ մեթոդը խորը ուսուցումն է: Այս հոդվածի նպատակն է գտնել Արհեստական բանականության վրա հիմնված լավագույն լուծումը հայերեն ձեռագիր տառերի ճանաչման համար: Քանի որ հայերենն այնքան էլ տարածված չէ, և դրա օգտագործումը սահմանափակ է, այդ իսկ պատճառով չկան արհեստական բանականության վրա հիմնված լավ մոդելներ կամ գործիքներ, որոնք կճանաչեն հայերեն ձեռագիրը:

Այս աշխատանքում օգտագործվել են խորը ուսուցման ամենահայտնի, արդյունավետ և ժամանակակից մեթոդները: Օգտագործվել է Mashtots տվյալների բազան, որը ստեղծվել է վերջերս Հայաստանի ազգային պոլիտեխնիկական համալսարանի «Ավտոմատացման համակարգեր և մոդելավորում» հետազոտական լաբորատորիայում: