A Systematic Mapping Study of Convolutional Neural Networks for Computer Vision and Other Applications

Gurupungav Narayanan
16CO114, Computer Engineering
National Institute Of Technology Karnataka
gurupungavn@gmail.com

Nihal Haneeg
16CO128, Computer Engineering
National Institute Of Technology Karnataka
nihalh55@gmail.com

***Abstract*—Deep Learning models which employ convolutional architectures have been found to be highly effective in image recognition and computer vision applications. Convolutional Neural Networks are now the primary architectures being implemented for computer vision applications like Optical Character Recognition and Image Classification. In this paper, we perform a systematic mapping study of Convolutional Neural Networks in order to explore existing material on different applications and implementations of Convolutional Neural Networks.**

I. Introduction

In the past decade, machine learning has evolved to become one of the most important fields of research. With large datasets now easily available to the public, increasing developments are being made in Machine Learning and Artificial Intelligence. Conventional Machine Learning models, which worked efficiently for small scale and simple applications, were unable to work as efficient when it came to Computer Vision algorithms. Basing the learning algorithm on features designed prior to training turned out to be extremely complex and unintuitive. This deficiency in conventional algorithms led to the development of deep learning algorithms. Deep learning models usually consist of multiple layers of non-linearity, where for every layer the image is transformed and abstract features are discovered. Deep learning models generally work on the image directly, taking the image as input. Conventional Neural Networks worked very efficiently and produced exceptional results as compared to conventional models. The aim of this paper is to study existing research and publications on the various applications of Convolutional Neural Networks. We attempt to gain insight into state-of-the-art implementations of Convolutional Neural Networks and distinguish potential new directions for research. The 131 papers selected have been classified by their field of interest, learning paradigm used, research type and contribution type. The second section describes the procedure used to analyze the publications. The third section answers the research questions which were defined in the second section. We, finally present our conclusions in section four.

II. Mapping Study

The objective of a systematic mapping study is to provide an insightful overview of a research field. A systematic mapping study builds a classification scheme and structure for a field of interest. The frequency of publications under that defined scheme is analyzed to produce results. A sharp increase in the number of papers published in a research field implies that the field is maturing. Multiple aspects of the scheme can also be investigated to produce more concrete answers to specific research questions.

The process thats being employed to perform the Mapping Study is the one thats proposed by Peterson et. al.[1] The process suggests the use of diagrams and visual representation to provide a summary of its results. The first step in process is to define the Research Questions which concretely helps us define the the scope of the research. The second step is to conduct the actual search by using search strings on scientific databases or by manually browsing through conference papers and journals. This search should, of course, be driven by the Research Questions defined and the search strings based on the structure of the classification. This is followed by the actual screening of the papers, to collect relevant papers and classify them based on keywords and the abstracts. The final step is to extract the data from the screened papers and complete the mapping process.

A. Research Questions and Search Queries

The aim of this study is to analyze various publications regarding convolutional neural networks and the various fields its applied in. We laid down Research Questions, shown in Table 1, to reflect that aim. The digital library used to search for publications is IEEE-Xplore, the IEEE Digital Library of publications. To perform more complex and specific searches, we defined search strings using various Boolean operators defined in Table 2. Since the screening step would discern the relevancy of papers, the search strings should be defined such that the search includes as many publications that are even remotely related to the field in question. In the screening phase, we defined an inclusion and exclusion scheme for deciding a publications relevance (or irrelevance).

Table 1: Research Questions

No.	Research Questions	Motivation
RQ1	In what fields are Convolutional Neural Networks employed in?	This question highlights the main aim of this paper and helps us to understand where recent studies are focusing on.
RQ2	What type of learning paradigm was used to train the neural network?	This question elucidates the kind of data that the networks were used to train on and the nature of the learning.
RQ3	What kinds of research have been proposed?	This question helps us understand the various research studies taking place in this area of study.
RQ4	What are the various contributions that have been made?	This question helps us understand the various contributions made in this area of study.

Table 2: Search String

No.	Search String	No. of Parts
SS1	((”Abstract”:convolutional neural networks AND ”Abstract”:computer vision) AND ”Abstract”:image processing)	86
SS2	((”Abstract”:convolutional neural networks NOT image NOT visual NOT Audio NOT sound))	28
SS3	((”Abstract”:convolutional neural networks AND ”Abstract”:applications) AND ”Abstract”:Audio)	18

The scheme is as follows:

1) Inclusion Scheme:
   - All books, papers, presentations and likewise material which provide information on systematic studies performed in the area of CNN in computer vision and other applications.
   - Recent studies will be considered if repetition of material is seen.
2) Exclusion Scheme:
   - Materials which have CNN just in the introductory paragraphs and/or in the abstracts.
   - Studies that do not show any useful or relevant information on our area of study.

The search strings defined fulfill at least one of the above defined schemes characteristics.

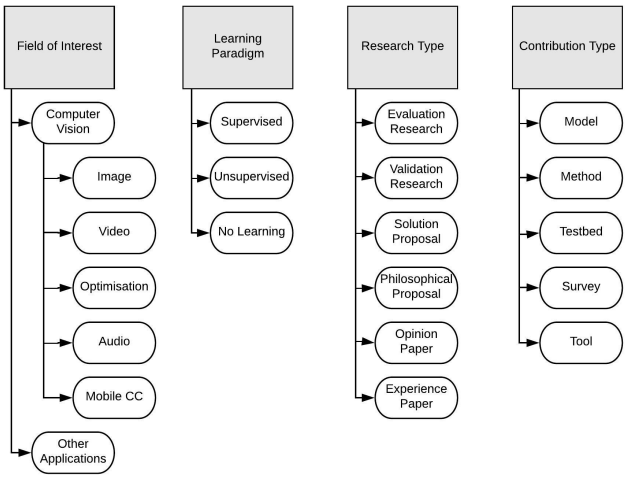

Fig. 1. Classification Scheme

B. Classification Scheme

Classification schemes are rated on the basis of a set of quality attributes [1]:

1) Orthogonality. There are clear boundaries between categories, which makes it easy to classify.
2) Defined based on existing literature. The taxonomy/classification is created on the basis of an exhaustive analysis of existing literature in the field.
3) Based on the terminology used in literature. The taxonomy uses terms that are used in existing literature
4) Complete. No categories are missing, thus allowing existing articles to be classified.
5) Accepted. The community accepts and knows the classification/taxonomy

The classification scheme used is presented in Figure 1. The first two classes are with respect to the field that we are trying to map. The keywords used in the classification scheme is very closely related to the Research Questions defined. This allows us to easily derive answers to those questions after classification. The field of interests under Computer Vision were derived after some tactful keywording by Nishani et. al. [2]. The research type and contribution type classifications being employed are suggested by Wierenga et. al. [3]. To classify the research papers into the following classes, we keyworded the abstracts of the publications. If the abstracts did not provide enough information, the Introduction and Conclusion sections were also keyworded.

III. Results

This section answers the research questions defined in Section II. Answering the research questions elucidates various facets of the field of interest. The results of the Systematic Mapping Study are contained in the following sections.

A. RQ1: In what fields are Convolutional Neural Networks employed in?

Table 3: Number of papers by Field of Interest

Field of Interest	Number of Papers	Percentage
Image	73	55.3
Audio	18	13.6
Video	13	9.8
Optimization	3	2.3
Others	25	18.9

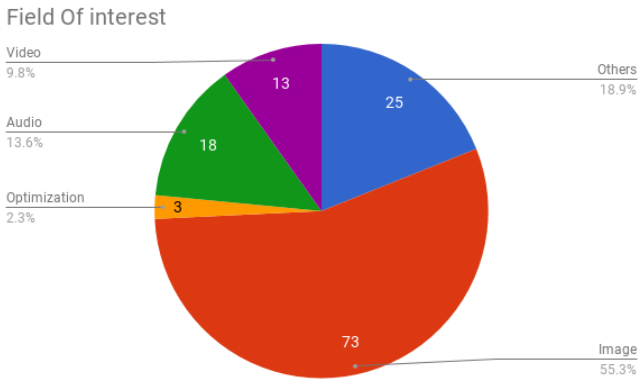

Fig. 2. Pie chart of Table 3

This question deals with the various fields of interests that is dealt by the papers. Table 3 shows our recorded observations from the scanning phase and Fig 2 visualizes that data into a pie chart. The figure above indicates that Convolutional Neural Networks are primarily used in computer vision applications.

55.3% of the papers used Convolutional Networks of some architecture to process images. Papers that used Convolutional Neural Networks for classifying images, like [4] which classified vegetables into categories or [5] which classified advertisement images, were very common.

13.6% of the papers dealt with some form of audio processing. Most of these papers used CNNs for speech recognition and sentiment analysis. This included papers like [6] which attempted to discern the emotion of the speaker using CNNs and [7] which used a multilevel CNN for detecting the cries of babies in a domestic environment.

9.8% of the papers used CNNs for video processing. Video processing was primarily used for action recognition and obstacle recognition in robotic applications. Papers like [8] and [9] demonstrate these applications.

A fair number of papers 18.9% used CNNs for applications that were not for Computer Vision or other related spheres. Many of such papers involved Natural Language Processing or other forms of analyzing text data, like [10] which used CNNs for Chinese text sentiment analysis and [11] which performed a sentiment analysis on Thai twitter data.

The data analysis suggests that Convolutional Neural Networks are primarily used for computer vision applications but other potential fields of effective application are continually being explored.

B. RQ2: What type of learning paradigm was used to train the neural network?

Table 4: Number of papers by Learning Paradigm

Learning Paradigm	Number of Papers	Percentage
Supervised Learning	112	84.8
Unsupervised Learning	20	15.2

This question is with regard to the learning paradigm used to train the neural network. Table 4 shows our recorded observations from the scanning phase. The table above indicates that Supervised Learning techniques are primarily used in training neural networks.

84.8% of the papers used Supervised Learning techniques to train their neural network models. Papers that used this technique for training their neural networks, most of which are in the area of image processing and computer vision, like [4] which classified vegetables into categories or [5] which classified advertisement images, were very common.

The rest of the papers which constitute to around 15.2% of the papers, used Unsupervised Learning for training the neural networks. For instance, paper [12] unsupervised learning technique was used to allow the model to learn to model visual objects in a way similar to the hand-designed features of a human being.

The data analysis suggests that Supervised Learning is primarily used by data scientists for computer vision and other applications but other potential fields of other effective training methods are continually being explored.

C. RQ3: What kinds of research have been proposed?

Table 5: Number of papers by Research Type

Research Type	Number of Papers	Percentage
Evaluation	43	32.6
Experience	11	8.3
Opinion	4	3.0
Solution	28	21.2
Validation	46	34.9

We derived the results of Table 5 by examining the research type of the chosen papers. Majority of the papers (34.9%) were Validation research. This basically means that solutions were validated in various simulation and controlled test scenarios. [13] uses a dataset of over 60000 images to recognize complex events from static images by fusing deep channels.

A significant portion of the published papers also contributed to evaluation research. Here, solutions were implemented and tested in real world environments. [7] which evaluated a model for detecting baby cries, also built a simple mobile based application which implemented the paper's results.

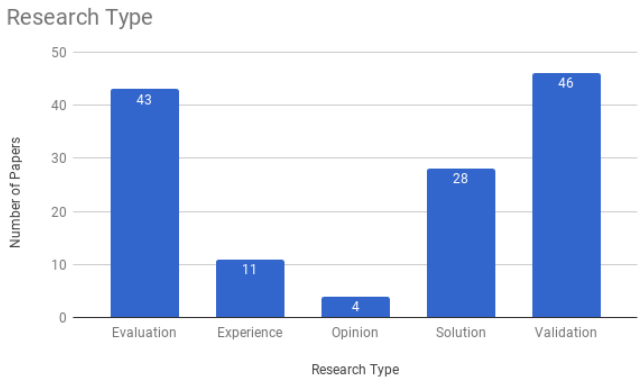

Fig. 3. Bar Graph of Table 5

21.2% of the papers proposed solutions that have not been evaluated in real world scenarios. This includes papers like [14] which proposes CNNs to be employed for real time object detection for unmanned aerial vehicles.

Very few papers were based on experience (8.3%) and opinions (3.0%). These mostly included papers that reviewed or described other publications.

D. RQ4: What are the various contributions that have been made?

Table 6: Number of papers by Contribution Type

Contribution Type	Number of Papers	Percentage
Method	66	50
Model	52	39.4
Testbed	1	0.8
Tool	13	9.8

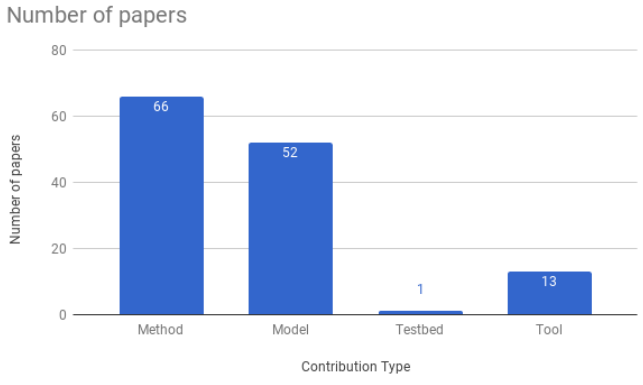

Fig. 4. Bar Graph of Table 6

39.4% of the papers propose new models. The models propose new frameworks, architectures, kernels etc to solve some problems more effectively. For instance, a new framework is proposed by [15] for human decomposing a human image into body regions. The CNN is employed to construct the relation between the input human image and the structure outputs for human parsing. [16] proposes a weakly supervised CNN framework called MILCNN to solve the insufficient training data problem.

A majority of the papers 50% propose novel methods. Methods proposed are generally new implementation techniques or learning algorithms. For example, [17] combines conventional CNNs with thermal RGB-D-T face recognition concepts to implement highly accurate facial recognition.

The remaining papers dealt with tools (9.8 %) and testbeds (0.8 %).

IV. Limitations of the Study

Any Study is always limited by the available resources, the bias of the researcher and the possible judgmental errors made unbeknownst to the researcher. There may be some threats to the validity of the study done.

- Conclusion validity. This threat arises due to incorrect data extraction or missing papers. The research criteria defined by the researcher might be subjective, biased or underfitting. This paper defines its classification scheme based on the research questions defined. In this context, the limitation is alleviated for this particular mapping study.
- Construct validity. In this case, the problem may be related to the identification of primary papers. It is probable that the search string may require different keywords or that the sources we used to collect papers may not be enough. In this case we have used one main Digital library: IEEE Xplore Digital Library. In the future the collection of papers can be done from other sources as well; for e.g. ACM or Springer.
- Internal validity. Data extraction and analysis might result in some unclear results and wrong conclusions. Since extraction is done only based on the abstract, this may limit the validity of the study.
- External validity. These kinds of threats are related to the generality of the results. Since this study is focuses only on a specific field and its facets defined by the research questions, this particular threat is irrelevant to this mapping study.

V. Conclusion

In this paper, we perform a systematic mapping study of Convolutional Neural Networks for Computer Vision and Other Applications. We reviewed existing literature with respect to various applications, implementations and frameworks of Convolutional Neural Networks and extracted data from them to gain a general understanding of the scope of the field. The data analysis concluded that:

1) Convolutional Neural Networks are primarily used in Computer Vision applications (79%) but other fields of applications are continually being explored.
2) Most models based on Convolutional Neural Networks use Supervised Learning (85%).

3) Majority of the papers published are usually Validation or Evaluation research based on existing solutions that employ Convolutional Neural Networks.
4) 50% of the papers propose new methods for Convolutional Neural Networks and 39.4% propose new frameworks and architectures. This suggests that various usecases of Convolutional Neural Networks are still being explored and identified.

Hence, the interest in the field is as high as it could be and it is very relevant to upcoming technologies and developments.

REFERENCES

[1] K. Petersen, R. Feldt, S. Mujtaba and M. Mattsson, 2008. Systematic Mapping Studies in Software Engineering.In Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering. Italy, 68-77.

[2] Eralda Nishani and Betim io. 2017. A Systematic Mapping Study of Computer Vision Approaches based on Deep Learning and Neural Network. In Proceedings of the 8th Balkan Conference in Informatics (BCI '17). ACM, New York, NY, USA, Article 10, 8 pages. DOI: https://doi.org/10.1145/3136273.3136301

[3] N. Maiden, N. Mead, C. Rolland and R. Wieringa. 2005. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requirements Engineering. Springer-Verlag 11, 1, 102-107. DOI:http://dx.doi.org/10.1007/s00766-0050021-6

[4] Y. Sakai, T. Oda, M. Ikeda and L. Barolli, "A Vegetable Category Recognition System Using Deep Neural Network", 2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), Fukuoka, 2016, pp. 189-192. doi: 10.1109/IMIS.2016.84

[5] A. T. Vo, H. S. Tran and T. H. Le, "Advertisement image classification using convolutional neural network," 2017 9th International Conference on Knowledge and Systems Engineering (KSE), Hue, 2017, pp. 197-202. doi: 10.1109/KSE.2017.8119458

[6] W. Lim, D. Jang and T. Lee, "Speech emotion recognition using convolutional and Recurrent Neural Networks," 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, 2016, pp. 1-4. doi: 10.1109/APSIPA.2016.7820699

[7] Y. Lavner, R. Cohen, D. Ruinskiy and H. Ijzerman, "Baby cry detection in domestic environment using deep learning," 2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE), Eilat, 2016, pp. 1-5. doi: 10.1109/ICSEE.2016.7806117

[8] J. Lee, J. Wang, D. Crandall, S. abanovi and G. Fox, "Real-Time, Cloud-Based Object Detection for Unmanned Aerial Vehicles," 2017 First IEEE International Conference on Robotic Computing (IRC), Taichung, 2017, pp. 36-43. doi: 10.1109/IRC.2017.77

[9] S. Kumra and C. Kanan, "Robotic grasp detection using deep convolutional neural networks," 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, 2017, pp. 769-776. doi: 10.1109/IROS.2017.8202237

[10] Kecong Xiao, Zishuai Zhang and Jun Wu, "Chinese text sentiment analysis based on improved Convolutional Neural Networks," 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2016, pp. 922-926.

[11] P. Vateekul and T. Koomsubha, "A study of sentiment analysis using deep learning techniques on Thai Twitter data," 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, 2016, pp. 1-6. doi: 10.1109/JCSSE.2016.7748849

[12] M. Y. Liu, A. Mallya, O. Tuzel and X. Chen, "Unsupervised network pretraining via encoding human design," 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, 2016, pp. 1-9. doi: 10.1109/WACV.2016.7477698

[13] Y. Xiong, D. Lin and X. Tang, "Recognize Complex Events from Static Images by Fusing Deep Channels", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, 1600-1609. doi: 10.1109/CVPR.2015.7298768

[14] J. Lee, J. Wang, D. Crandall, S. abanovi and G. Fox, "Real-Time, Cloud-Based Object Detection for Unmanned Aerial Vehicles," 2017 First IEEE International Conference on Robotic Computing (IRC), Taichung, 2017, pp. 36-43. doi: 10.1109/IRC.2017.77

[15] X. Liang, S. Liu, X. Shen, J. Yang, L. Liu, J. Dong, L. Lin and S. Yan, " Deep Human Parsing with Active Template Regression", IEEE Transactions On Pattern Analysis And Machine Intelligence, 37, 2, 2402-2414. DOI: http://dx.doi.org/10.1007/10.1109/TPAMI.2015.2408360

[16] M. Sun, T. X. Han, Ming-Chang Liu and A. Khodayari-Rostamabad, "Multiple Instance Learning Convolutional Neural Networks for object recognition," 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, 2016, pp. 3270-3275. doi: 10.1109/ICPR.2016.7900139

[17] M. O. Simn et al., "Improved RGB-D-T based face recognition," in IET Biometrics, vol. 5, no. 4, pp. 297-303, 12 2016. doi: 10.1049/iet-bmt.2015.0057