Transfer Learning Techniques for Image Recognition: A Systematic Literature Review

Tariq Mahmood [1] Aleenah Khan [2] Salma Tajuddin [3]

Abstract

Deep learning is a rapidly expanding research area focusing on the use of more extended (deep) and varied neural network architectures to solve more complicated problems than traditional multi-layer perceptrons. Transfer learning is a more recent off-shoot of deep learning which focuses on using information from one machine learning task in another related task. It has primarily seen applications in image classification, for instance, when information used to recognize/classify a bicycle can be used to classify a motorcycle. In a rapidly evolving research space, it is important to summarize the research applications of different deep learning off-shoots. In this regard, this paper presents the first systematic literature review particularly targeting applications of transfer learning to image recognition. We follow the standard methodology and categorize papers on the basis of more critical KPIs. Our core finding is that this particular domain is a hot area of research these days, and most applications are related to pre-trained models learnt from convolution neural network and applied to another convolution network. Also, transfer learning has lead to significant improvements in accuracy and efficiency and facilitation, as compared to learning deep models or other machine learning approaches from scratch. From our results, we propose several future directions of research.

Keyword: Transfer Learning, Image Recognition, Image Classification, Systematic Literature Review, Convolution Neural Networks

1 Introduction

Deep learning [1], [2], [3], [4] involves the use of larger and more robust artificial neural network models, as compare to traditional single-layer or multi-layer perceptrons [5]. The word 'deep' mostly signals the use of more hidden layers containing a larger number of hidden neurons, along with novel, diverse architectures for instance, convolution neural networks, recurrent neural networks (LSTM, GRU), restricted Boltzmann machines, deep belief networks, and generative models [2]. Deep learning has successfully solved many complicated machine learning problems, for instance, image recognition, image captioning, machine translation, natural language processing, and automatic speech recognition [6], [7], [8]. Notwithstanding this, deep learning always has had a problem of efficiency with initial training times running into hours. The rapid evolution and application of GPU technology has catered for this problem to a certain extent [9], [10], efficiency still remains

[1] *Institute of Business Administration, Karachi | tmahmood@iba.edu.pk*
[2] *Institute of Business Administration, Karachi | aleenahkhan@iba.edu.pk*
[3] *Institute of Business Administration, Karachi | stajuddin@iba.edu.pk*

a major problem. Considering the complexity of deep learning problems, there is also a need to make learning easier. If we train an LSTM to recognize English speech patterns of American users, we would want to use this knowledge for recognizing speech of British users also because both of them come from the same distribution of English speaking users. Knowledge from a convolution network model that recognizes cars can be used in another convolution network that is learning to recognize trucks, because both belong to the same distribution of motor vehicles.

Transfer learning [11], [12], [13], [14], [15] is a machine learning technology which uses models (architecture with hyper parameters) from one learning problem in another related learning problem. The first research paper appeared in 1992 [15] and the field has seen significant advances since then, particularly in deep learning. In this paper, we perform a systematic literature review of the applications of transfer learning to image classification or recognition. In particular, we are motivated by the importance of image recognition to the biomedical domain. Image classification is a critical application area of deep learning and there is relatively substantial literature on its transfer learning applications which needs to be reviewed. Although there have been several reviews on transfer learning applications [16], [11], this is the first review giving a drill-down into image recognition. We adopt the standard reviewing procedure by identifying keywords, search queries, filtration of retrieved papers, and then creating appropriate labels to classify the results. Due to some limitation of time, we are able to present the most recent results till the time of conducting this survey, i.e., from January 2017 till April 2017. Even with this limited sample, we were able to extract useful results and propose concrete future directions of research.

2 Background Knowledge

A Convolution Neural Networks

Convolution Neural Networks [17], [18], [19] (CNNs) are feed-forward deep neural networks best suited to solve visual imagery learning problems, e.g., image classification and recognition. They are famous because they eliminate the need to exact image features

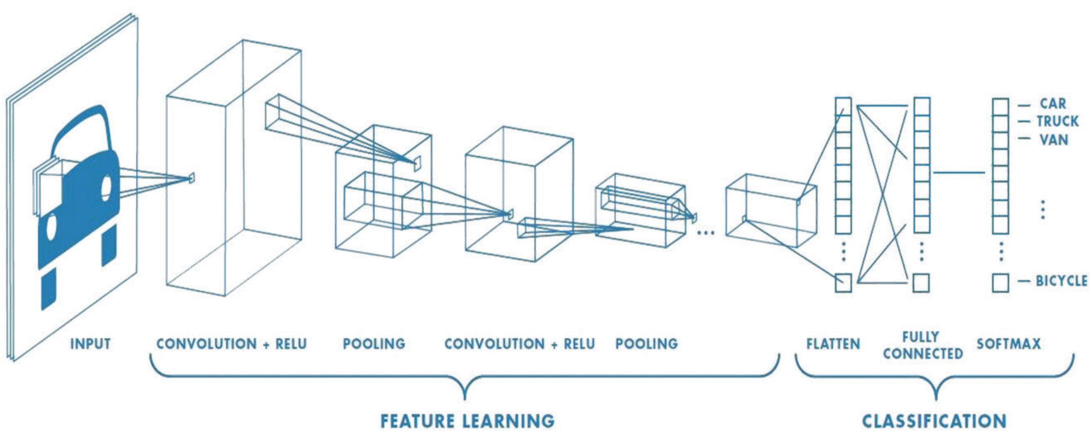

Figure 1: Generalized Architecture of Convolution Neural Network

Manually and also facilitate transfer learning. Figure 1[4] shows the basic architecture of a CNN. We can imagine all the layers in feature learning and classification sections as being stacked together from bottom (feature learning) to top (classification) in 3D, having width, height and depth. The depth is typically 3, corresponding to RGB color bands. The width and height correspond to the same dimensions of an input image in 2D. In feature learning, the convolution layer applies a particular filter on the input image to extract one or more features and the RELU activation function non-linearizes the result. Each neuron in the convolution layer might only recognize a particular feature or part of an image (e.g., an edge, circle, color shade etc.) in its receptive field, i.e., a given neuron is only connected to a small part of the image and takes no input from the remaining image parts. This latter concept is taken from research in neuroscience in how animals and humans recognize objects. The job of the pooling layer is to summarize the non-linearized convolution filters, e.g., by taking maximum or average. Typically, a large number of convolution + RELU + Pooling triples are applied, in order to automatically extract as many features as possible from the input image and summarize them. In the classification phase, the output from the final pooling layer is flattened out into a vector which is fed to a traditional fully-connected MLP. In the output, a soft max activation function is applied to assign probabilities to output, depending on the type of output being trained on.

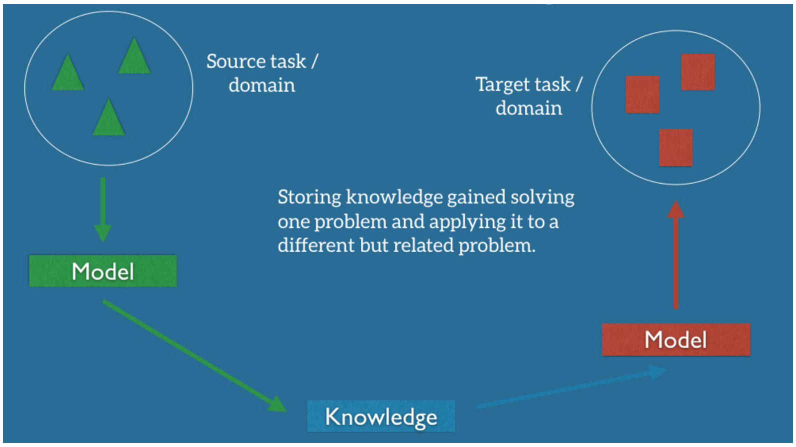

Figure 2: Concept of Transfer Learning

B Transfer Learning

Transfer learning is a technique of machine learning in which a model trained on one problem is re-purposed on another related problem. This transfer of knowledge typically improves generalization in the other setting and hence can also be considered a type of optimization of the second task. Transfer learning is generally related to multi-task learning and concept drift, and cannot be considered only related to deep learning.

However, it is prominent in deep learning because of extensive computational resources required to train and learn deep models. The requirement here is that model features learned in the initial problem should be generic in nature. Specifically, the source network is trained on a

[4] *Adapter Network HTML from mathwork.com/discovery/convolutional-neural-netwrok.html*

baseline task with a baseline dataset. Its learned features are then re-purposed, or transferred, to the target network for learning a target task with a target dataset. This situation is shown in Figure 2[5]. The transfer will work if the features used are applicable to both the source and target tasks, and not just specific to one of them. This inductive transfer considerably minimizes the hypotheses space for searching the solution for target task.

There are two approaches to deep transfer learning: 1) model development, and 2) using pre-trained models. In the first case, we first select a related source problem which could be less complicated than the target one, based on relatively similar inputs and outputs. We then learn the source task and ensure that a significant amount of feature selection has occurred, and then re-purpose this model as the starting point for the target task. Here, the model can be used partially or completely and is then finally tuned for the target. In the second case, we reuse a model pre-trained on complicated deep problems, which are provided by research organizations and academia. Again, partial or complete model can be used, which is then finally tuned to the target problem. This latter approach is more commonly used in deep learning. Nowadays, transfer learning papers are published in acclaimed conference related to data science and machine learning, for instance, ICDM, KDD, ICML, AAAI, NIPS, ECML, and NIPS.

C Formalizing Transfer Learning

Transfer learning can be formalized by answering three questions:

1) What needs to be transferred?: Knowledge can either be applicable to only a single task or it can be applicable across multiple tasks. We need to determine the partition of knowledge which will be transferred from source to target task.

2) When should the transfer take place?: We need to determine the exact situations in which transfer should happen and those in which transfer learning is not applicable. In the latter case, forcefully transferring knowledge through brute force can lead to derogatory performance, also called a "negative transfer".

3) How should the transfer be done?: Transfer learning can occur either from one type of machine learning problem into another one. We need to determine the exact categories and then the most appropriate algorithm for this type of transfer.

1) *Transfer Learning with Visual Imagery:* Visual imagery problems are most common in deep learning, particularly related to object recognition. The input is an image in colored or gray-scale mode, and the task is to recognize objects in the image, for instance, motor vehicles, animals or people. More than one object (class) needs to be detected for a given object type, for instance, cars, trucks and motorcycles in motor vehicle object type, and cats, dogs and rabbits in animal object type. For such problems, we can use models pre-trained for more challenging image recognition tasks such as the standard Image net Large Scale Visual Recognition Challenge (image-net.org/challenges/LSVRC/2016/index) challenging global researchers on a 1000-class problem. Models pre-trained on this problem are provided by several research

[5] *Adapter from ruder.io/transfer-learning/*

groups, for instance, the visual geometry group at university of Oxford[6], Microsoft's Resent CNN model[7] and Google's inception model[8]. More such models can be found at Caffe's Model Zoo[9]. These pre-trained models are more useful in the earlier layers of the target task, and more specific target dataset features are useful in later layers.

More formally, a source image recognition problem with domain D consists of feature space F and a marginal probability distribution $P(G)$ over feature space with $G = f1, f2,\ldots,fn \epsilon F$ with G being the total set of images used in training. Given $D = F; P(G)$, problem τ has labeled space γ with the conditional probability distribution $P(Y—G)$ learned from training data consisting of pairs $fi \epsilon G$ and yi $\epsilon \gamma$. In our case, γ is the set of image object type labels, e.g., car, bus, and truck. Now suppose we have a source domain Ds, source problem τ_s, target domain D_t and target problem τ_t. The aim of transfer learning is to learn the distribution P(Yt | Xt) in Dt with knowledge acquired from Ds and τ_s where $D_s \neq D_t$ and $\tau_s \neq \tau_t$. This formulation gives rise to the following four possible scenarios:

$F_s \neq F_t$: The feature space of the source and target tasks are different from each other, e.g., the input images represent different object types (cars and truck)

$P(G_s) \neq P(G_t)$: The marginal distributions of source and target domains are different, e.g., distribution of features of cars and trucks will be different

${}_s \neq {}_t$: The label spaces is different for each domain, e.g., images are assigned different types of labels in the target space (this case doesn't occur with a high probability)

$P(Y_s \mid G_s) \neq P(Y_t \mid G_t)$: The conditional probability distributions of the source and target tasks are different from each other, e.g., if the labeling is unbalanced in both cases with respect to the class distribution.

One method of doing transfer learning is to execute it using simulations (to avoid training of expensive hardware) and the reuse the results in the real world target task. In this case, case 2 above is applicable and not case 1 as the feature space is the same in both simulation and real-world. The simulation technique is extremely useful for complicated tasks like training self-driving car; Fig 3 shows such a simulator from udacity [20]. Another pertinent concept is that of domain adaptation, in which the source and target data could belong to different domains. This happens when we are forced to use a pre-trained model on something which is not exactly what we want. For instance, one may use a model pre-trained to recognize regular bike in shopping stores to learn to recognize mountain bikes being driven on mountain roads and tracks.

[6] *robots.ox.ac.uk/vgg/research/very_deep/*
[7] *github.com/KaimingHe/deep-residual-networks*
[8] *github.com/tensorflow/models/tree/master/research/inception*
[9] *github.com/BVLC/caffe/wiki/Model-Zoo*

Figure 3: Udacity Self-Driving Car Simulator

3 Systematic Literature Review

A systematic literature review (SLR) is a repeatable research method which critically analyzes multiple research studies in parallel regarding a particular research domain, with the intention of answering one or more research questions according to a structured methodology [21]. In this paper, we conduct the first SLR on applications of transfer learning to image recognition which is the first of its type to the best of our knowledge. We chose an SLR because we want to investigate about this research domain rather than find some particular solution for it. Our research questions are formulated as follows:

RQ1: How important are transfer learning applications to image recognition for the research community? We will answer this by considering the number of related publications.
RQ2: What is the impact of transfer learning on image recognition? We will answer this by considering the results of experiments in the publications along with other related data.
RQ3: What are the future research directions of transfer learning applications to image recognition? We will answer this by initially analyzing our review results and then deriving the potential future directions.

We limit our searched relevant papers to those provided by developed queries ()given below) from renowned computer science digital sources, i.e., IEEE, ACM, Google Scholar, Springer, Elsevier, and Wiley. Duplicates were likely due to the presence of Google Scholar in the list sources which can reiterate results from any and all other sources list. We removed these duplicates before any analysis.

A Keyword Identification and Query Formation

Based on our research domain and scope of review, we identified keywords and generated search queries based on those keywords. Although most keywords can be combined with each other, convolution neural networks is a subset of deep learning and therefore these terms may not be used in conjunction with one another. We kept such relationships into account when forming queries. We identified ten types of logically possible conjunctions of our keywords, which are shown in Table I. It is important to note that these conjucntions represent the context on what we searched and do not represent the exact query that we used. While querying, we applied various types of rephrasing to get better results, especially in the context of inadequately sized data set. This particular concept was difficult to formulate precisely and we searched it using various alternates. The results documented were the maximum of all result counts, instead of summing up in case of repetitions. In some case when different and possibly relevant results were observed, we formed a combination of results using the Mendeley citation management tool[10].

Table I: Context of our Search Queries

No.	Search Query
1	Deep Learning Techniques and Image Classification
2	Transfer Learning Techniques and Image Classification
3	Convolution Neural Networks and Image Classification
4	Convolution Neural Networks and Transfer Learning
5	Image Classification and Transfer Learning
6	Handling Inadequate Sized Data Classification of Images using Deep learning
7	Handling Inadequate Sized Data Classification of Images using Transfer learning
8	Handling Less Data Classification of Images with Convolutional Neural networks
9	Image Classification using Convolutional Neural Networks and Transfer Learning
10	Image Classification with Transfer Learning and CNN (small data sets)

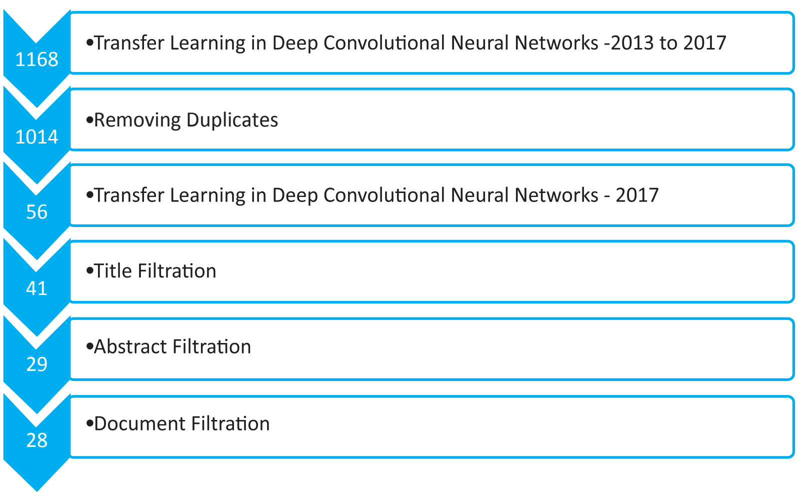

Figure 4: Article Filtration Process

[10] *www.mendeley.com*

B Article Filtration Process

On executing the search queries, we retrieved a total of 1168 relevant research publications from our digital sources. These papers were published between 2013 and 2017 (inclusive) and specifically dealt with applications of transfer learning to convolution neural networks. We then adopted an article filtration process shown in Figure 4. Removing duplicates using Mendeley gave us 1014 articles. Based on certain constraints, we considered this a large quantity of literature to review, so we decided to focus initially on the most recent publications in 2017. This left us with 56 articles up to the time of this review. We then used a three-pronged approach: we first filtered for title (remaining 41 articles), then for abstract (remaining 29 articles) and finally the body of the publication (remaining 28 articles).

C Analytics of Retrieved Articles

The number of publications per year is shown in Figure 5. As we can see, there is an almost exponential increase in the number of articles related to transfer learning from 2013-2016. It is expected that the total number of articles in 2017 will be greater than 300, but to keep our scope limited, we only consider 28 publications up to the time of writing this survey (January 2017 - April 2017).

Moreover, as can be seen in Figure 6, most of these 28 articles have been published in conference proceedings. There are no technical reports and just 1 book chapter and thesis along with 2 journals. One reason for this is the fast pace at which the transfer learning field is growing which makes it more convenient to publish in a conference with shorter review times as compared to other types.

Also, the distribution of our 28 articles with respect to digital sources (Figure 7) shows that a large number of articles were retrieved from Google Scholar, followed by IEEE, arXiv, IEEE, Elsevier and finally ACM.

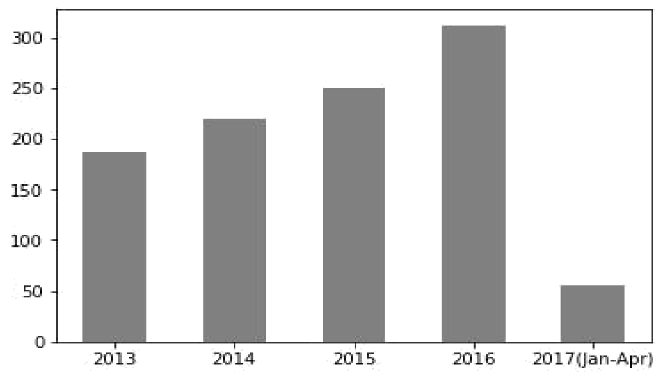

Figure 5: Annual Distribution of Articles from 2013-2017

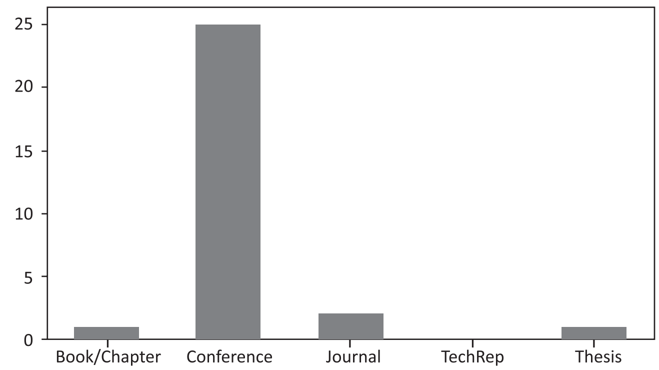

Figure 6: Distribution of Articles with respect to Type

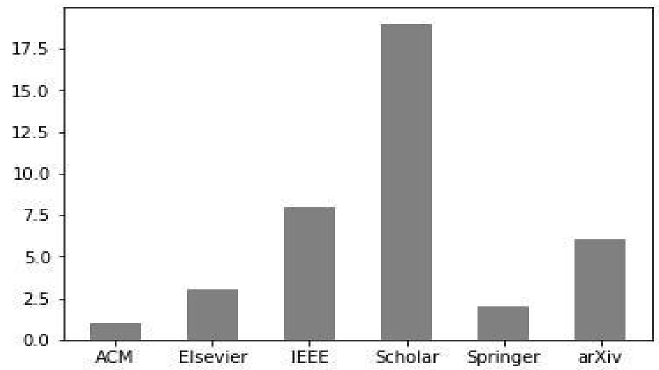

Figure 7: Distribution of Articles with respect to Digital Sources

The usage of pooling functions in our 28 articles is shown in Figure 9. Mostly, the well-known max pooling is used, followed by average pooling. However, a large majority of articles haven't mentioned their pooling function and a single paper hasn't mentioned their pooling function. The activation function distribution is shown in Figure 10. As can be expected, mostly ReLU has been used, followed by Softmax. A single paper has used Sigmoid while the remaining have not mentioned their activation function details.

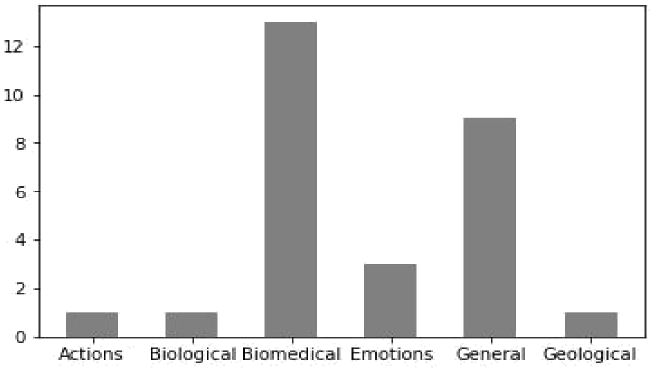

Figure 8: The Domain Type of Research in 28 Selected Articles

Table 2: Datasets Used in Selected 28 Articles

GBM	KIRC	NIH AREDS	BCDR-F03	Indian Pines
CIFAR10	MRSCVOC2007	USPS	Salinas Valley	Pavia University
MNIST	COIL20	CMU-PIE	PACS (Abdonminal Ultrasounds)	ISBI 2016 Challenge Dataset
Office	Caltech256	BUAA VIS-NIR	Chalearn-LAP First Impressions	Driver Monitoring
LifeCLEF2015	DLCST	COPDGENE1	RSD	aPY
COPDGENE2	Frederickshavn	OxfordDogs120	GENKI	AM-FED
OxfordFollowers102	MITIndoor67	Duke OCT	NJUD	STEREO
APR	EmotiW2015	EmotiW2016	AwA	UCM
CK+	MMI	RECOLA	DES	NLPR

The data sets used in our 28 papers are shown in Table II. These are classification datasets. The number of categories varies from only four entries to a set of hundreds of classes. The size of the data varies from 200 to around 50,000. This is all possible with good results due to transfer learning. Some experiments gradually increase the training set size to observe the impact of transfer learning and its ability to perform well with even a small set of labeled images. Also, the type of domain research with respect to our articles is shown in Figure 8. We see that almost 50% of articles are from Bio-Medical domain followed by general datasets. Three articles deal with human emotion recognition, while individual contributions come from human actions, land type and plant recognition domains.

In Table III, we create a classification label for each selected article and also show the overall performance results which are extracted from the paper. We found two papers dealing with unsupervised transfer learning [22], [23], and two which have presented a robust framework for learning metrics or hyper parameters through transfer learning [24], [25]. The majority of classification is understandably related to application of transfer learning in deep convolution neural networks [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42] to solve problems in diverse application domains shown in Figure 8. Two papers have addressed the use of transfer learning in non-deep convolution networks [43], [44], while the remaining have used non-CNN methods to enhance image classification performance through transfer learning [45], [46], [47], [48], [49].

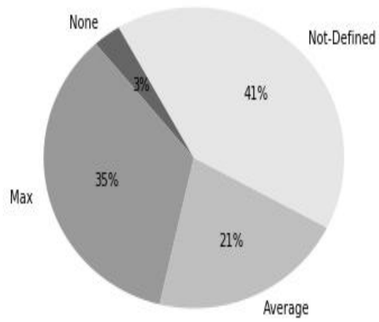

Figure 9: Distribution of Pooling Technique Usage

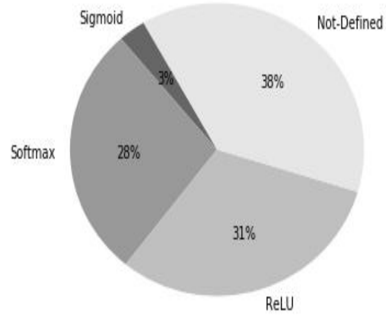

Figure 10: Distribution of Activation Function Usage

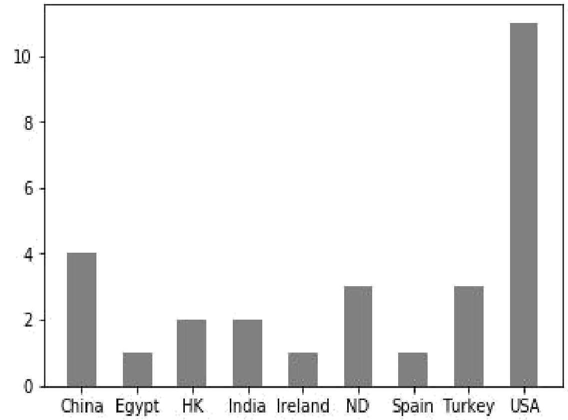

Figure 11: Distribution of Papers with respect to Country of First Author - ND=Netherlands, HK=Hong Kong

Regarding the performance results, most articles have recorded accuracy measure on their selected datasets. In Table III, we have shown the average performance across all the selected datasets for the sake of clarity and understanding. We see that most accuracies for deep CNN based transfer learning applications are satisfactory (90%). Apparently, such an architecture seems to be best validated for transfer learning support in image recognition. The other recorded measures are the Area Under the Curve (AUC) [43], Root Mean Squared Error (RMSE) [31], overall error in classification [39], and an improvement in performance [44], [47]. However, the performance in these papers is also seen to be satisfactory. In essence, it is apparent that the application of transfer learning does have a substantial effect on improving the performance, although researchers have preferred to stick more with deeper CNN based models due to their recent success, e.g., LeNet, AlexNet, GoogleNet, ResNet, VGGNet and CaffeNet.

Finally, we extracted the country of the first author for each of our selected articles to understand the global locations where more research in our domain is concentrated. These results are shown in Figure 11. Majority of the papers have been published by American authors (11), followed by China (4), Netherlands (3) and Turkey (3). Other countries with even lesser publications include Ireland, Spain, India and Egypt. We didn't have publications from Australia, South America and Africa.

4 Future Research Directions

Our research (although performed on a limited scale) has identified several crucial details. Based on this, we are able to suggest the following future research directions:

Application to Security Domain: The most frequent application of our selected papers is in the biomedical domain. However, we couldn't find any application in security or cyber security realm where face, gesture and movement recognition become extremely crucial. Applications in this domain need to increase substantially.

Increase in Journal Publications: As compared to the conference publications, journal articles are almost infrequent. This clearly shows that the transfer learning field is expanding at a rapid pace and hence is more suited to shorter review times of conference. However, in order to lay deeper scientific foundations of the field, the number of journals also needs to increase.

Performance with Different Pooling and Activation Functions: In our papers, researchers have gone with the use of standard max pooling and ReLU activation function. However, we didn't find any article offering a comprehensive comparison of different pooling techniques and activation functions. We deem it necessary considering the substantial performance improvements offered by transfer learning.

Application with Other Machine Learning Approaches: We have seen significant number of applications using deep convolution networks. However, the performance with other architectures and algorithms has been as satisfactory. There needs to be hence more research in these directions: either exploration of non-CNN architectures and algorithms or im-proving transfer learning through other machine learning approaches coupled with CNN architectures. These combinations can also be used with other deep learning approaches, for instance, recurrent neural networks for tasks such as image captioning or image description generation [2].

Experimentation with CNN hyper parameters: In [50], the authors establish a robust taxonomy of CNN hyper parameters which have generated to the huge success of CNNs in the recent past. There needs to be a comprehensive research work which experiments with all such parameters with one or more standard transfer learning approaches. Such hyper parameters include different types of convolution layers (tiled, transposed, dilated, network in network, inception module), loss functions (hinge, softmax, contrastive, triplet, KL divergence), optimization (data augmentation, weight initialization, stochastic gradient descent, batch normalization, shortcut connections), and faster processing methods (fast fourier transforms, structured transforms, low precession, weight compression and sparse convolution).

5 Answering the Research Questions

We now reiterate and answer our research questions as follows:

RQ1: How important are transfer learning applications to image recognition for the research community? Answer: Transfer learning applications are considered highly beneficial to image recognition domain. In the period 2013-2016, the numbers of related publications have increased at an exponential rate and hence, this is a hot research area these days. Mostly, the research has focused on healthcare and biomedical domains primarily to automate recognition across clinical laboratory test images, e.g., MRI, CT Scan etc.

RQ2: What is the impact of transfer learning on image recognition? Answer: The impact of transfer learning on image recognition is substantial. In our review, we considered 28 publications for only four months in 2017. We still found that satisfactory performance is achieved in all these articles, and we consider this to be better than the typical deep CNN or MLP performance based on our experience.

RQ3: What are the future research directions of transfer learning applications to image recognition? Answer: Although this is a hot area of research, still much needs to be done. We have listed our proposed directions in Section IV. Mostly transfer learning focuses on using knowledge from pre-trained CNNs in another CNN but we particularly need to investigate the effect of transferring knowledge learnt from other machine learning models and approaches (e.g. SVMs).

Table 3: Classification of Selected 28 Papers and their Performance

Paper Ref.	Performance	Classification of Paper
[22]	Accuracy: >90%	Unsupervised TL
[26]	Accuracy: >84%	Deep CNN based TL
[27]	Accuracy: 77%	Deep CNN based TL
[43]	AUC: 0.85	CNN Training + Linear Discriminate Analysis TL
[24]	Accuracy: >90%	Transfer Metric Learning Framework
[25]	Accuracy: >75%	Transfer Metric Learning Framework
[44]	65% Improvement in Performance	Improving CNN based TL for MRI
[28]	Accuracy: >80%	Deep CNN based TL
[45]	Accuracy: >90%	Logistic Classifier Training + Weighted Logistic Classifier TL
[29]	Accuracy: >90%	Deep CNN based TL
[42]	Accuracy>93%	Deep CNN based TL
[46]	Accuracy: >85%	Improving TL in Unshared Feature Space
[30]	Accuracy: >80%	Deep CNN based TL
[31]	RMSE: 5.6	Deep CNN based TL
[32]	Accuracy: >70%	Deep CNN based TL
[47]	Improved Performance	Facial Recognition Transfer Learning
[48]	Accuracy: >90%	Active Transfer Learning Approach

[49]	Accuracy: >90%	SVM Based Transfer Learning
[33]	Accuracy: >85%	Deep CNN based TL
[23]	Accuracy: >90%	Unsupervised TL
[34]	Accuracy: >72%	Deep CNN based TL
[35]	Accuracy: >90%	Deep CNN based TL
[36]	Accuracy: >90%	Deep CNN based TL
[37]	Accuracy: >90%	Deep CNN based TL
[38]	Accuracy: >84%	Deep CNN based TL
[39]	Error: 7.1%	Deep CNN based TL
[40]	Accuracy: >90%	Deep CNN based TL
[41]	Accuracy: >73%	Deep CNN based TL

6 Conclusion

This paper has presented a systematic literature review on applications of transfer learning approach to image recognition. We have followed the standard procedure for conducting such a review. Due to several constraints, we reviewed publications only in period January 2017 - April 2017. Even in this limited time frame, we retrieved 28 relevant articles and analyzed them thoroughly over the more standard metrics. Results show that transfer learning brings a significant boost in efficiency and accuracy to image recognition, particularly using convolution neural networks. Based on current domains, we motivate its application to other domains such as cyber security and also indicate the need to thoroughly present an experimentation with all transfer learning procedures for image recognition with convolution neural networks. Our future research work is to review the 1000-odd papers from 2013 to 2017 and to validate our results across this time period. We also plan to conduct an experimentation of standard transfer learning methods with changes in CNN hyper parameters as suggested in our research directions in Section IV.

Acknowledgment

The authors would like to thank Usman Ali (PhD Scholar at Institute of Business Administration) for his valuable feedback on this paper.

References

[1] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in Computer Vision – ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 818–833.

[2] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning.

[3] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in NIPS Deep Learning and Representation Learning Workshop, 2015. [Online]. Available: http://arxiv.org/abs/1503.02531

[4] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition." CoRR, vol. abs/1310.1531, 2013.

[5] E. Alpaydin, Machine Learning: The New AI.

[6] L. Deng and D. Yu, Deep Learning: Methods and Applications. Hanover, MA, USA: Now Publishers Inc., 2014.

[7] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," Mechanical Systems and Signal Processing, vol. 107, pp. 241 – 265, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327017306064

[8] D. Yu and L. Deng, "Deep learning and its applications to signal and information processing [exploratory dsp]," IEEE Signal Processing Magazine, vol. 28, no. 1, pp. 145–154, Jan 2011.

[9] Y. Cai and S. See, GPU Computing and Applications, 1st ed. Springer Publishing Company, Incorporated, 2016.

[10] J. Nickolls and W. J. Dally, "The gpu computing era," IEEE Micro, vol. 30, no. 2, pp. 56–69, Mar. 2010. [Online]. Available: http://dx.doi.org/10.1109/MM.2010.41

[11] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. on Knowl. and Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010. [Online]. Available: http://dx.doi.org/10.1109/TKDE.2009.191

[12] B. Tan, Y. Song, E. Zhong, and Q. Yang, "Transitive transfer learning," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '15. New York, NY, USA: ACM, 2015, pp. 1155–1164. [Online]. Available: http://doi.acm.org/10.1145/2783258.2783295

[13] T. Galanti, L. Wolf, and T. Hazan, "A theoretical framework for deep transfer learning," Information and Inference: A Journal of the IMA, vol. 5, no. 2, pp. 159–209, 2016. [Online]. Available: + http://dx.doi.org/10.1093/imaiai/iaw008

[14] L. Yang, S. Hanneke, and J. Carbonell, "A theory of transfer learning with applications to active learning," Mach. Learn., vol. 90, no. 2, pp. 161–189, Feb. 2013. [Online]. Available: http://dx.doi.org/10.1007/s10994-012-5310-y

[15] L. Y. Pratt, "Discriminability-based transfer between neural networks," in Advances in Neural Information Processing Systems 5, [NIPS Conference]. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993, pp. 204–211. [Online]. Available: http://dl.acm.org/citation.cfm?id=645753.668046

[16] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," Journal of Big Data, vol. 3, no. 1, p. 9, May 2016. [Online]. Available: https://doi.org/10.1186/s40537-016-0043-6

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 1097–1105. [Online]. Available: http://dl.acm.org/citation.cfm?id=2999134.2999257

[18] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in Shape, Contour and Grouping in Computer Vision. London, UK, UK: Springer-Verlag, 1999, pp. 319–. [Online]. Available: http://dl.acm.org/citation.cfm?id=646469.691875

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 770–778.

[20] D. Etherington, "Udacity open sources its self-driving car simulator for anyone to use," https://beta.techcrunch.com/2017/02/08/udacity-open-sources-its-self-driving-car-simulator-for-anyone-to-use/, 2017, [Online; accessed 9-February-2017].

[21] B. Kitchenham, "Procedures for performing systematic reviews," Department of Computer Science, Keele University, UK, Keele University. Technical Report TR/SE-0401, 2004.

[22] H. Chang, J. Han, C. Zhong, A. Snijders, and J.-H. Mao, "Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2017. [Online]. Available:https://doi.org/10.1109/tpami.2017.2656884

[23] R. Kemker and C. Kanan, "Self-taught feature learning for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 5, pp. 2693–2705, may 2017. [Online]. Available: https://doi.org/10.1109/tgrs.2017.2651639

[24] Z. Ding and Y. Fu, "Robust transfer metric learning for image classification," IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 660–670, feb 2017. [Online]. Available: https://doi.org/10.1109/tip.2016.2631887

[25] Y. Xu, S. J. Pan, H. Xiong, Q. Wu, R. Luo, H. Min, and H. Song, "A unified framework for metric transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 6, pp. 1158–1171, jun 2017. [Online]. Available: https://doi.org/10.1109/tkde.2017.2669193

[26] P. Burlina, K. D. Pacheco, N. Joshi, D. E. Freund, and N. M. Bressler, "Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis," Computers in Biology and Medicine, vol. 82, pp. 80–86, mar 2017. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2017.01.018

[27] K. H. Cha, L. M. Hadjiiski, H.-P. Chan, R. K. Samala, R. H. Cohan, E. M. Caoili, C. Paramagul, A. Alva, and A. Z. Weizer, "Bladder cancer treatment response assessment using deep learning in CT with transfer learning," in Medical Imaging 2017: Computer-Aided Diagnosis, S. G. Armato and N. A. Petrick, Eds. SPIE, mar 2017. [Online]. Available: https://doi.org/10.1117/12.2254977

[28] M. M. Ghazi, B. Yanikoglu, and E. Aptoula, "Plant identification using deep neural networks via optimization of transfer learning parameters," Neurocomputing, vol. 235, pp. 228–235, apr 2017. [Online]. Available: https://doi.org/10.1016/j.neucom.2017.01.018

[29] W. Ge and Y. Yu, "Borrowing treasures from the wealthy: Deep transfer learning through selective joint fine-tuning," CoRR, vol. abs/1702.08690, 2017. [Online]. Available: http://arxiv.org/abs/1702.08690

[30] H. Kaya, F. Gurpınar,¨ and A. A. Salah, "Video-based emotion recognition in the wild using deep transfer learning and score fusion," Image and Vision Computing, vol. 65, pp. 66–75, sep 2017. [Online]. Available: https://doi.org/10.1016/j.imavis.2017.01.012

[31] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 10, no. 5, pp. 1963–1974, may 2017. [Online]. Available: https://doi.org/10.1109/jstars.2017.2655112

[32] "RGB-D salient object detection based on discriminative cross-modal transfer learning," CoRR, vol. abs/1703.00122, 2017, withdrawn. [Online]. Available: http://arxiv.org/abs/1703.00122

[33] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," in Proceedings of the 5th International Conference on Bioinformatics and Computational Biology, ser. ICBCB '17. New York, NY, USA: ACM, 2017, pp. 59–62. [Online]. Available: http://doi.acm.org/10.1145/3035012.3035022

[34] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, jan 2017. [Online]. Available: https://doi.org/10.1038/nature21056

[35] P. M. Cheng and H. S. Malhi, "Transfer learning with convolutional neural networks for classification of abdominal ultrasound images," Journal of Digital Imaging, vol. 30, no. 2, pp. 234–243, nov 2016. [Online]. Available: https://doi.org/10.1007/s10278-016-9929-2

[36] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," CoRR, vol. abs/1707.07012, 2017. [Online]. Available: http://arxiv.org/abs/1707.07012

[37] M. S. Elmahdy, S. S. Abdeldayem, and I. A. Yassine, "Low quality dermal image classification using transfer learning," in 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE, 2017. [Online]. Available: https://doi.org/10.1109/bhi.2017.7897283

[38] A. Romero-Lopez, X. G. i Nieto, J. Burdick, and O. Marques, "Skin lesion classification from dermoscopic images using deep learning techniques," in Biomedical Engineering. ACTAPRESS, 2017. [Online]. Available: https://doi.org/10.2316/p.2017.852-053

[39] M. Turan, Y. Almalioglu, E. Konukoglu, and M. Sitti, "A deep learning based 6 degree-of-freedom localization method for endoscopic capsule robots," CoRR, vol. abs/1705.05435, 2017. [Online]. Available: http://arxiv.org/abs/1705.05435

[40] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land–cover classification of high-resolution imagery,"

IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 4, pp. 549–553, apr 2017. [Online]. Available: https://doi.org/10.1109/lgrs.2017.2657778

[41] J. Lemley, S. Bazrafkan, and P. Corcoran, "Transfer learning of temporal information for driver action classification," in 28th Modern Artificial Intelligence and Cognitive Science Conference, 04 2017.

[42] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," Biomedical Optics Express, vol. 8, no. 2, p. 579, jan 2017. [Online]. Available: https://doi.org/10.1364/boe.8.000579

[43] B. Q. Huynh, N. Antropova, and M. L. Giger, "Comparison of breast DCE-MRI contrast time points for predicting response to neoadjuvant chemotherapy using deep convolutional neural network features with transfer learning," in Medical Imaging 2017: Computer-Aided Diagnosis, S. G. Armato and N. A. Petrick, Eds. SPIE, mar 2017. [Online]. Available: https://doi.org/10.1117/12.2255316

[44] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. G. Guttmann, F.-E. de Leeuw, C. M. Tempany, B. van Ginneken, A. Fedorov, P. Abolmaesumi, B. Platel, and W. M. Wells, "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in Lecture Notes in Computer Science. Springer International Publishing, 2017, pp. 516–524. [Online]. Available: https://doi.org/10.1007/978-3-319-66179-7 59

[45] V. Cheplygina, I. P. Pena, J. H. Pedersen, D. A. Lynch, L. Sorensen, and M. de Bruijne, "Transfer learning for multi-center classification of chronic obstructive pulmonary disease," IEEE Journal of Biomedical and Health Informatics, pp. 1–1, 2017. [online]. Available: https://doi.org/10.1109/jbhi.2017.2769800

[46] W. Chang, Y. Wu, H. Liu, and Y. Yang, "Cross-domain kernel induction for transfer learning," in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA., 2017, pp. 1763–1769.

[47] A. Anuse, N. Deshmukh, and V. Vyas, "Transfer subspace learning model for face recognition at a distance," vol. 9, pp. 27–32, 01 2017.

[48] Y. Guo, G. Ding, Y. Wang, and X. Jin, "Active learning with cross-class knowledge transfer," in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, ser. AAAI'16. AAAI Press, 2016, pp. 1624–1630. [Online]. Available: http://dl.acm.org/citation.cfm?id=3016100.3016126

[49] I. Becht, "Individualized smile detection using transfer learning," Master's thesis, Utrecht University, 2017.

[50] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," Pattern Recognition, vol. 77, pp. 354–377, may 2018. [Online]. Available: https://doi.org/10.1016/j.patcog.2017.10.013