

Neural Networks for Tool Image Classification

Boas Bamberger¹, Oliver Erlenkaemper² and Fabian Wolf³

Abstract—Augmented reality solutions for field workers are an emerging market. [1] An augmented reality solution for field workers requires software perceiving the environment of field workers. A subtask of that perception is the classification of different tools. We seek to determine the best-performing neural network for tool image classification. Furthermore, we introduce a novel dataset for tool image classification (TIC Dataset). To determine the best-performing neural network for tool image classification we train and evaluate neural networks on the TIC Dataset. We select the neural networks for the experiment based on a literature review which we conducted. We found that, in general, not one particular but several neural networks are suitable for tool image classification, especially neural networks using convolutional layers and skip connections.

I. INTRODUCTION

Augmented reality solutions for field workers are an emerging market. [1] An augmented reality solution for field workers requires software perceiving the environment of field workers. A subtask of that perception is the classification of different tools. For example, a software displaying step-by-step instructions in the field of vision of a field worker needs to distinguish a screwdriver from a wrench when telling the field worker to tighten the screw with the wrench lying on the ground next to him instead of the screwdriver in his hand. This task is called tool image classification. We seek to determine the best-performing neural network for tool image classification. Furthermore, we introduce a novel dataset for tool image classification the TIC Dataset. To determine the best-performing neural network for tool image classification we train and evaluate neural networks on the TIC Dataset. We select the neural networks for the experiment based on a literature review which we conducted. Our time and resources were limited. For this reason, we exclusively train the neural networks supervised without auxiliaries. On that account, we exclude metalearning, especially neural architecture search, non-neural machine learning models, other computer vision tasks, unsupervised learning, semi-supervised learning, transfer learning, adversarial training, data augmentation, input normalization, weight decay, and multi-task learning.

II. RELATED WORK

We seek to determine the best-performing neural network for tool image classification. To the best of our knowledge our paper is the first paper on tool image classification. Tool image classification is a subtask of image classification which

is an advanced field and mostly solved by neural networks. Neural networks achieve accuracies of 96.08% – 98.66% on CIFAR, MNIST, SVHN, STL, and 88.61% on the ImageNet Dataset [2], [3] which comprises 1000 classes and 3.2 million images [4]. Especially convolutional neural networks [5], residual neural networks [6], inception networks [7] and dense neural networks [8] have advanced the field.

III. METHOD

We determine the best-performing neural network for tool image classification in the course of an experiment on the TIC Dataset. Optimally, we would conduct a neural network architecture search and hyperparameter search for the experiment, but since our computational resources are limited we select the neural networks for the experiment based on a literature review which we conducted.

A. Dataset Construction

The TIC Dataset is a dataset for tool image classification. Thus, we created images consisting of exactly one tool of different classes. We chose the classes based on the tools which were freely available to us. The resulting classes are drill, hammer, pliers, saw, screwdriver and wrench. In real world scenarios, tools are presented from various angles and with various backgrounds. Due to this, the created images are taken from arbitrarily chosen angles and with arbitrarily chosen backgrounds. To create an image we placed a tool in an arbitrarily chosen position and took a series of close-up images. During the series, the camera was moved around to create arbitrary angles. For the next series, the tool, the background, and/or the position of the tool was changed. The resulting dataset is comprised of 20,400 images divided in 6 balanced classes.

B. Experiment

We selected the neural networks based on a literature review. The literature review was conducted according to the method of [9] to examine the state of the art of image classification in regard to neural network architectures. For each architecture we selected one neural network for the experiment. Since the best-performing neural network on the ImageNet Dataset for each architecture uses more RAM than available to our limited hardware [2], [10], [11], we selected the neural network from the paper originally proposing the architecture. For the experiment, we split the TIC Dataset 60%/20%/20% into training development and test dataset. Training and hyperparameter tuning was executed exactly as in the paper originally proposing the neural network, except

¹Scientific Supervisor bamberger@uni-mannheim.de

²Supervisor oliver.erlenkaemper@honeyswell.com

³Author s172298@student.dhbw-mannheim.de

for data augmentation, weight decay and dropout which we excluded in this paper, see Section I. Furthermore, we needed to adapt the batchsize to the size of the available GPU RAM. We evaluated the performance of each neural network on the test dataset in accuracy.

IV. EVALUATION

The accuracy for each neural network on the TIC Dataset is reported in Table I. For comparison we also report the original accuracy for each neural network on the ImageNet Dataset. Among these neural networks, DenseNet-264 performs best for the TIC Dataset. ResNet-152, ResNeXt-101, and DenseNet-264 perform rather similar. Therefore, we conclude that several neural networks are suited for tool image classification. EfficientNet-B7 achieved an accuracy of 16.67%, meaning it did not learn at all.

TABLE I
EXPERIMENT RESULTS

Neural Network	TIC Dataset	ImageNet
VGG-19	72.40	76.3 [12]
ResNet-152	92.89	78.57 [6]
ResNeXt-101	94.66	79.6 [13]
DenseNet-264	97.45	77.85 [8]
EfficientNet-B7	16.67	84.4 [14]

V. DISCUSSION

We found that, in general, several neural networks are suitable for tool image classification. This is in accordance with the state of the art of image classification, as image classification leaderboards show that different neural networks perform rather similarly for different image classification tasks. [2]

Especially, we found, that ResNet-152, ResNeXt-101, and DenseNet-264 are suitable for tool image classification. This also was to be expected since, ResNet-152, ResNeXt-101, and DenseNet-264 are suitable for image classification in general. [13], [6], [8] Despite differing in structure, each of these neural networks uses convolutional layers and skip connections. [13], [6], [8] This might be the cause of the similar performance. In comparison to these neural networks, VGG-19 performs less well. VGG-19 does not use skip connections. [12] Thus, we conclude, not using these connections might be the cause of the lower performance of VGG-19.

In comparison to the ImageNet Dataset ResNet-152, ResNeXt-101, and DenseNet-264 perform better on the TIC Dataset, this might be due to the different complexity of the datasets. ImageNet Dataset has 1000 classes, while the TIC Dataset has 6 classes. Furthermore the classes are completely different. Accordingly the complexity of classification is different for each class.

EfficientNet-B7 did not learn at all. This might be, because, due to the limited GPU RAM, we were only able to train EfficientNet-B7 with a batch size of 1. A batch size of 1 causes the loss function to fluctuate heavily. This

impairs convergence to the optimum and thus learning.[15] On that account, the accuracy of EfficientNet-B7 reported in Section IV cannot be interpreted as the performance of EfficientNet-B7 for tool image classification, but simply as that EfficientNet-B7 was not able to learn in the course of the experiment. We exclusively train the neural networks supervised without auxiliaries. On that account, it is likely, that even higher accuracies can be achieved using the excluded techniques in addition to our approach.

VI. CONCLUSION

We conclude that, for tool image classification convolutional neural networks using skip connections are suited. For industry aiming to create augmented reality solutions for field workers, our implementation can be used as basic module of software perceiving the environment of field workers. For future work, we propose to improve our work with more performant hardware, additional training data and to implement metalearning, semi-supervised learning, transfer learning, adversarial training, data augmentation, input normalization, weight decay, and multi-task learning, since these techniques were shown to improve performance. [16], [17], [18] Furthermore, we propose future work on the implementation of a holistic augmented reality solution for field workers, i.e. further computer vision tasks, a software framework, and a hardware framework for implementation.

REFERENCES

- [1] McKinsey, The coming evolution of field operations (2019).
- [2] P. Foret, A. Kleiner, H. Mobahi, B. Neyshabur, Sharpness-aware minimization for efficiently improving generalization (2020).
- [3] H. M. D. Kabir, M. Abdar, S. M. J. Jalali, A. Khosravi, A. F. Atiya, S. Nahavandi, D. Srinivasan, Spinalnet: Deep neural network with gradual input (2020).
- [4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database, in: CVPR, 2009, pp. 248–255.
- [5] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature.
- [6] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: CVPR, 2016, pp. 770–778.
- [7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: CVPR, 2015, pp. 1–9.
- [8] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: CVPR, 2017, pp. 2261–2269.
- [9] J. Webster, R. T. Watson, Analyzing the past to prepare for the future: Writing a literature review, MIS Quarterly.
- [10] A. I. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, N. Houlsby, Large scale learning of general visual representations for transfer (2019).
- [11] H. Touvron, A. Vedaldi, M. Douze, H. Jegou, Fixing the train-test resolution discrepancy, in: NIPS, Curran Associates, Inc, 2019, pp. 8252–8262.
- [12] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR.
- [13] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: CVPR, 2017, pp. 5987–5995.
- [14] M. Tan, Q. Le V, Efficientnet: Rethinking model scaling for convolutional neural networks, in: ICML, 2019, pp. 6105–6114.
- [15] S. Ruder, An overview of gradient descent optimization algorithms (2016).
- [16] S. J. Pan, Q. Yang, A survey on transfer learning, TKDE.
- [17] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. J. Goodfellow, R. Fergus, Intriguing properties of neural networks, CoRR.
- [18] H. El-Amir, M. Hamdy, Deep Learning Pipeline: Building a Deep Learning Model with TensorFlow, 2020.