

Research Seminar in Data Science

Paper Reviews

Christoph Schmidl
s4226887
`c.schmidl@student.ru.nl`

June 6, 2018

Reviewed papers:

- **Human-level control through deep reinforcement learning** by Volodymyr Mnih et al. (Presented by Valentin Koch)
- **Deep Residual Learning for Image Recognition** by Kaiming He et al. (Presented by Stephan Dooper)

1 Human-level control through deep reinforcement learning

1.1 Summary

In his paper "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", Jun-Yan Zhu et al propose a method using Generative Adversarial Networks (GANs) in combination with adversarial and cycle consistency loss in order to perform image-to-image translation without the need of paired training data. The authors essentially try to map the distribution of one set of images X to another set of images Y in order to translate a certain style of one set to another. They do this with the help of GANs where one network generates an input using an image from X and try to transfer it to an image of Y . The same network structure is performed on the other side where another GAN tries to transfer the generated image X to Y back to X without any loss in information. This is called cycle consistency.

1.2 Evidence

The authors cite 65 different research papers and describe techniques which were used before and techniques which stay in direct competition to their proposed technique. Their proposed technique is based on GANs and their technique seems to be a solid improvement in the image-to-image translation domain. Their paper is therefore backed up by a sufficient amount of literature. The experiment and evaluation also seems legit. The evaluation is based on human perception by using AMT perceptual studies (Amazon Mechanical Turk) but also on automated techniques like using FCN score. They used a dataset for evaluation which were used before by other researchers and put their findings into perspective to prior research. Nicely done.

1.3 Strengths

The paper is written in a clear way and it was easy to follow. The evaluation process made a lot of sense to me because they are using human perception by using AMT perceptual studies (Amazon Mechanical Turk) but also on automated techniques like FCN score to distinguish fake generated images from real images. The proposed technique like combining GANs with the idea of cycle consistency loss seems simple but I personally think it is a rather great idea.

1.4 Weaknesses

I think the paper is overall writing really well. One thing that was not really clear to me was if there has to be a similar image mapping between set X and Y . In other words, does this technique work although if there is no matching image from one set to the other? Do the sets have to have one kind of bijective relationship so that it works correctly?

1.5 Evaluation

I guess this paper would get accepted to a conference or journal because it adds knowledge on top of already known techniques and proposes a novel approach for the image-to-image translation domain with regards to a lack of a paired image set.

1.6 Comments on the quality of the writing

The writing style was formal but also written in a clear way. It is one of those papers which I enjoyed reading and which made a lot of sense to me.

1.7 Queries for discussion

- The authors state that the proposed technique does not perform well when it comes to geometrical translations like translating a dog to a cat. Is there a new technique which solved this problem? What would be possible ideas to solve this?
- Can this approach using cycle consistency also be applied to other domains than image generation? Is it already applied to other domains?

2 Deep Residual Learning for Image Recognition

2.1 Summary

Kaiming He et al address the degradation problem by introducing a deep residual learning framework. The degradation problem occurs with increased network depth. When the network starts converging the accuracy gets saturated and then degrades rapidly. In contrast to a plain network architectures like VGG16, a deep residual network is able to maintain a relatively high accuracy even when the number of layers varies from 100 to 1000. Kaiming He et al hypothesize that it is easier to optimize the residual mapping than to optimize the original, unreferenced mapping. Residual learning is realized by so called "shortcut connections". They evaluate their results based on experiments with the imagenet and cifar-10 datasets which are about image classification and the PASCAL and MS COCO dataset which are about object detection and object segmentation. Residual networks seem to yield better performance on all datasets when the number of layers start increasing into the hundreds.

2.2 Evidence

The authors cite 49 different research papers to back up their paper. Among the different citations are well known researchers like C.M. Bishop, I.J. Goodfellow, G.E. Hinton, Y. LeCun and J. Schmidhuber. Out of these 49 papers, only 3 are also published by K. He himself. The papers seem to be well picked and show the problems of training networks with increased depths and already known solutions on how to tackle these problems. Based on the fact that none of the cited papers includes the name "residual", I assume that this paper is the first one which came up with residual network structures.

2.3 Strengths

The authors explain in detail what the main problem is when it comes to training very deep neural networks. The degradation effect seems to be of main interest and the residual network structure seems to be a novel approach. In the section about "shortcut connections" the authors also mention that there seems to be a similar approach to their shortcut connection implementation which has been developed during the same time span by Schmidhuber. Schmidhuber's idea is called "highway networks". The authors state out that the main differences between their idea and the one proposed by Schmidhuber. I guess that was done because Schmidhuber is known for claiming credit for ideas which seem similar to his own but differ slightly and is known for confronting researcher publicly during conferences to explain themselves.

2.4 Weaknesses

I could not find any real discussion, conclusion or future work section. This made it kind of frustrating for me to read because I wanted to start to follow the main three-step approach of reading research papers.

2.5 Evaluation

I assume that this paper was accepted for a conference or journal because it offers a novel idea which seems promising. Maintaining the same amount of hyperparameters as so called "plain networks" but being able to make the network way deeper at the same time and keeping a high accuracy seems to be a very tempting idea to try out. Given the fact that the authors also won different competitions with their approach is another reason why this paper is interesting. Well known datasets like imagenet, cifar-10 and coco have been chosen to validate the approach which makes it easy to reproduce.

2.6 Comments on the quality of the writing

The technical detail seems to be sound and it is well written when you overlook the fact that the discussion, conclusion and future work section is missing.

2.7 Queries for discussion

- The authors state that they did not use maxout/dropout for their networks with over 1000 layers although they assume that this may improve their results. Has there been further work on this assumption?