Zürcher Hochschule
für Angewandte Wissenschaften

School of
Engineering

InIT Institut für angewandte
Informationstechnologie

Project work (IT18tb_WIN)

Research-PW: Continual Deep Learning for Visual Recognition

Autoren	Warwara Panyushkina, panyuwar
Hauptbetreuung	Thilo Stadelmann, stdm
Nebenbetreuung	Helmut Grabner, grbn

Datum	23.12.2021

Zurich University
of Applied Sciences

School of
Engineering

DECLARATION OF ORIGINALITY

Project Work at the School of Engineering

By submitting this project work, the undersigned student confirm that this work is his/her own work and was written without the help of a third party. (Group works: the performance of the other group members are not considered as third party).

The student declares that all sources in the text (including Internet pages) and appendices have been correctly disclosed. This means that there has been no plagiarism, i.e. no sections of the project work have been partially or wholly taken from other texts and represented as the student's own work or included without being correctly referenced.

Any misconduct will be dealt with according to paragraphs 39 and 40 of the General Academic Regulations for Bachelor's and Master's Degree courses at the Zurich University of Applied Sciences (Rahmenprüfungsordnung ZHAW (RPO)) and subject to the provisions for disciplinary action stipulated in the University regulations.

City, Date: Hermetschwil 23.12.21

Signature: [signature]

The original signed and dated document (no copies) must be included after the title sheet in the ZHAW version of all project works submitted.

Zurich University
of Applied Sciences

Table of Contents

Note of Thanks

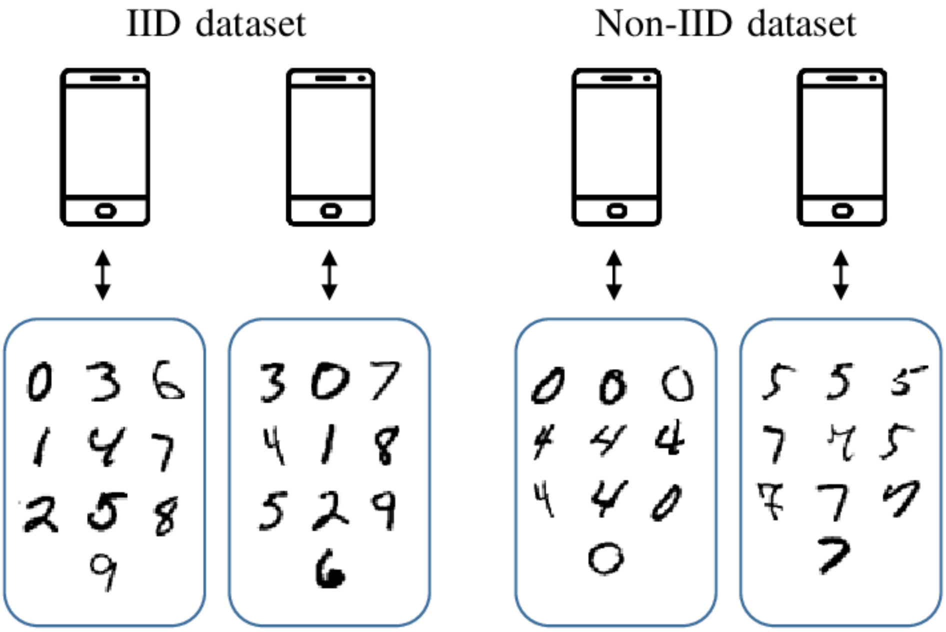
I would like to thank Thilo Stadelmann and Helmut Grabner for the opportunity to dive in into the research project and for their guidance.

Abstract

Since the coming of more continual data streams such as webcam, video streaming, robotic inputs and other, Continual Learning becomes more interesting topic of research. In this project I will try to introduce current research topics and to show viability for web-camera data. I will explain what kind of problem appeared during the project, their possible solutions. I try to explain the answer to why continual learning is a good approach for web-camera data by using theoretical knowledge I acquired during the research stage.

Introduction

In many real-world applications, the data we receive comes from a non-iid (non-independent or identical distributed) stream, meaning that if we receive an image of a car, we cannot guarantee that the next image will also be a car. Additionally, several restrictions may apply. Most common ones are limited storage space and security reasoning, where we are not able to keep the data indifferently long.



(Hellström et al., 2020)

We as humans cope with such tasks without noticing them. Since birth we learn things continuously and recall information when needed. We call this ability life-long learning, it is considered one of the most crucial abilities, which helps us develop more complicated skills throughout our lifetime (Nguyen et al., 2020). While we learn, we may forget some of the information over time, the process is gradual. This is normal in natural cognitive systems, where new information is almost never entirely forgotten. On the contrary in machine learning models, it can occur that the new object we learned, changes the weight distribution to a degree, where the previously learned objects won't be recognized anymore. This phenomenon is called Catastrophic Forgetting (French, 1999).

Related Works

Learning

Stephen Grossberg has started his work in 1960 on a systematic and effective approach to studying intelligent systems and deriving corresponding algorithms (Brna et al., 2019)(Grossberg, 1967). There are several interesting principles in his papers, most interesting are Stability-Plasticity Dilemma, Adaptive Resonance Theory (ART) (Grossberg, 2013)(Brito da Silva et al., 2019) and Knowledge Destillation (Furlanello et al., 2018).

Stability-Plasticity Dilemma

The Stability-Plasticity Dilemma is a well-known problem for neural networks in biological and artificial systems. For one the system has to have plasticity to adapt new knowledge, but also be stable enough not to forget existing one (Mermillod et al., 2013)(Grossberg, 2012).

Visual recognition and limitations of machine learning models

From 1958 onwards, David H. Hubel and Torsten Wiesel performed several experiments on different animals, showing that many neurons in visual cortex have a small local receptive field, meaning they react only to visual stimuli located in a limited region of the visual field (Géron, 2019, p. 446).

Convolutional Neural Networks

An important building Block of a Convolutional Neural Network (CNN) is the Convolutional Layer. An image cannot be thought of as a simple vector of input pixel values, because adjacency of pixels matters (Russel and Norvig, 2020). CNN are capable to successfully capture spatial and temporal dependencies in an image.

In a Convolutional Layer, neurons are not connected to all pixels, depending on the Layer it focuses on low-level features and assembles them into higher-level ones later on (Géron, 2019).

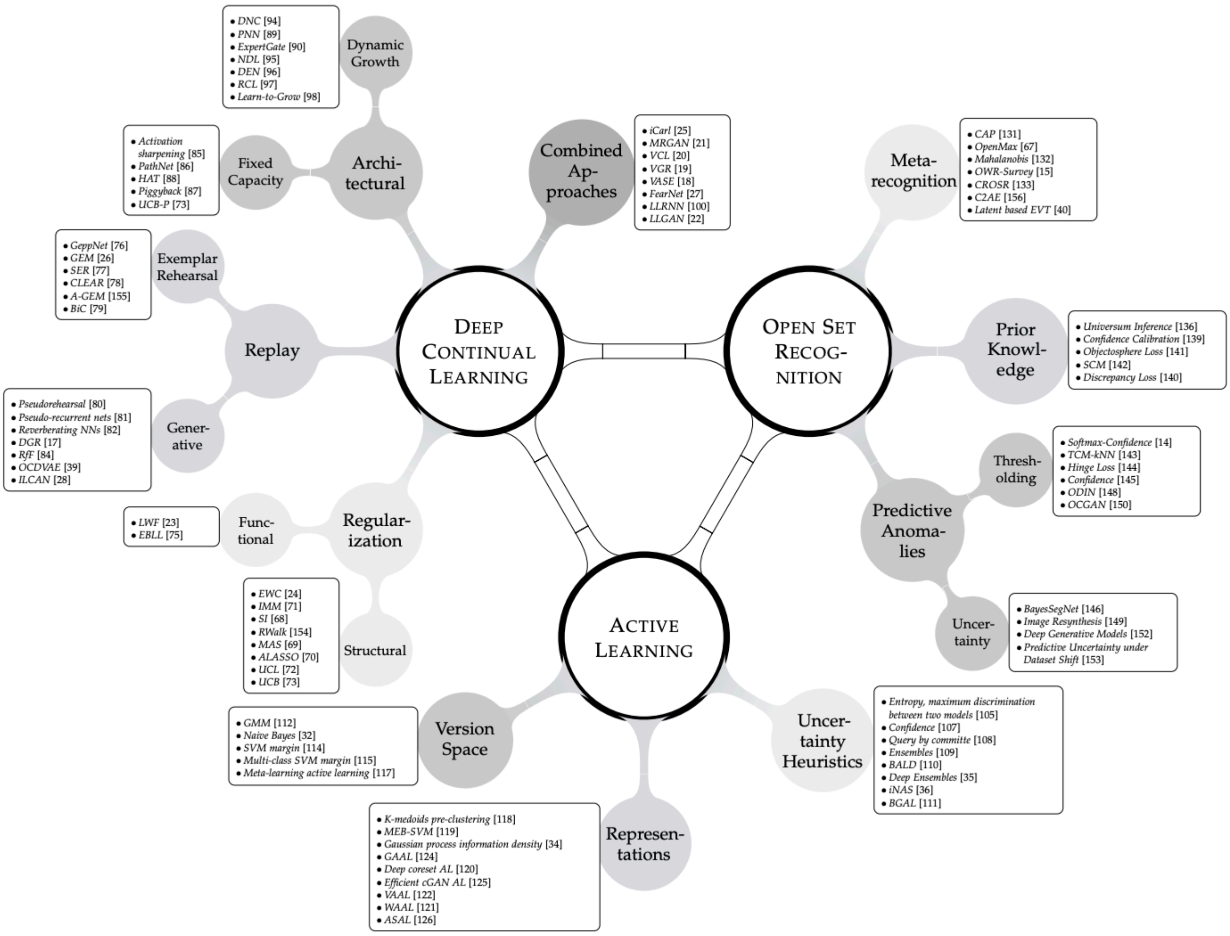
Catastrophic Forgetting

Catastrophic Forgetting happens in neural networks where data is forgotten abruptly on learning new information. The opposite is catastrophic remembering where the ability to discriminate between old and new inputs is reduced (Kaushik et al., 2021). While there are several studies on Catastrophic Forgetting (French, 1999)(Kirkpatrick et al., 2017)(Lee et al., 2018), many focus on external factors, but not why Catastrophic Forgetting happens internally (Nguyen et al., 2020).

Life-long learning / Continual learning (CL)

While humans can learn and adapt to changes in environment very fast, machine learning models often struggle with such task. In most cases the solution is to train them to a specific task or retrain when new information becomes available.

There many definitions for continual learning, most are sharing certain aspects. The stream of data is expected to be a non-iid stream (Mai et al., 2021). The stability-plasticity dilemma and catastrophic forgetting are central points. Currently many different methods are available in CL, a good overview is presented in (Lesort et al., 2019) and (Mundt et al., 2020).



(Mundt et al., 2020)

Some of these approaches can be attributed to several categories.

Problem

Many practical vision applications require learning new visual capabilities while maintaining performance on existing ones (Li and Hoiem, 2017). For the Webcam project of ZHAW I want to show whether Continual Learning is a better solution than retraining the model on new input. Currently we have incoming data from several Cameras around the world that are trained on a model with binary classifier. Users can adjust the model results to help it learn better. If the user relabelled the data, at night the model is retrained on this input. This method is inefficient, first the amount of data that is stored for retraining can grow very fast, secondly depending on data sensitivity we would either act in a grey area of computer security or cripple our model by removing data again.

I want to compare this approach with CL approach from Lifelong Machine Learning (Chen and Liu, 2016).

An Ideal Solution would remove the need for user Input, would be able to update the model without the need for retraining and be resistant to Catastrophic Forgetting.

Experiment

For the experiment I planned to compare the time and efficiency of the naïve model and a continual learning model (a naïve model is the one that is being retrained on new input). For the naïve approach a simple Convolutional Neural Network algorithm was used, that was adapted to the data from the ZHAW web-camera project (ajinkya98, 2021). For the continual version an example from "Learning without forgetting" (Li and Hoiem, 2017) (ngailapdi, 2021) and community code ("Papers with Code - Learning without Forgetting," n.d.) were used.

Other algorithms were considered, such as Hypernetworks (von Oswald et al., 2020) (Henning, 2021), different implementations of LWF ("Papers with Code - Learning without Forgetting," n.d.) and Sidetuning ("sidetuning," n.d.) (Sax et al., 2019). Due to time limitations it was decided against using them.

Results

During the project several problems appeared.
First, I had to switch from Keras to PyTorch in middle of project when I found the code implementations for continual solutions. This was followed by the problem with Cuda not being well supported by Nivida on MacOS, the main OS for the development. To fix this I tried port the environment to a singularity container to run it on the ZHAW GPU cluster. From there several combability issues arose which couldn't be solved without needing to recreate everything from scratch on a new linux machine, this was not possible due to time limitation.

Python environment

The Anaconda environment that was used on the MacOS for local development couldn't be transferred to the Linux system. Neither by hand nor by other means that were suggested, like using environment.yml or the package module ("Containerize a conda environment in a Singularity container," n.d.) ("hpc - Containerize a conda environment in a Singularity container," n.d.) ("Conda-Pack — conda-pack 0.6.0 documentation," n.d.) (*Conda-Pack*, 2021). This made it impossible to run it on the GPU Cluster without recreating it from scratch in the last weeks. The only possible solution would be to create a Dual-boot system on local computer and run it with correct configuration, since VirtualBox environments were providing similar errors.

Keras and Pytorch

Another problem was that as a starting point Keras was used and the switch to PyTorch during the project was a step back while trying to acquire know how. For the next project the improvements to be made, is to gather code examples before starting to develop even the simple approach. To see what at that moment is the most common approach.

Latex

The problem with using Latex was not having prior experience. The ZHAW template that was found, was from 2012 and had shown several bug in different Latex-Editors ("Overleaf, Online LaTeX Editor," n.d.)("TeXstudio - A LaTeX editor," n.d.)("Texmaker (free cross-platform latex editor)," n.d.) ranging from not compiling or updating parts of the document to not being able to compile at all.

Conclusion

The current papers on continual learning mostly focus on already prepared Datasets like MNIST or CIFAR-10 to train and benchmark their approaches. The real world doesn't have such ideal conditions. It would be interesting to see more approaches using real world data or a combination of different Datasets. Depending on Application and changes in the environment the variation in images can range from minimal to nothing comparable in the data that came before. This could falsify the algorithm to a certain degree.
Another point is the security concern, that is seldom considered in training an algorithm. A real-world application is often exposed to users who be it maliciously or not might input data that can produce unwanted results.
Additionally it would be interesting to combine more of CL Algorithms with Visualization of Catastrophic forgetting to better understand where it occur (Nguyen et al., 2020).

After better understanding the constrains and limitations of my development operation-system, I continue work on the project and finalize it.

Useful Tools

During the project some useful tools were found, such as Zotero, a helpful citation software ("Zotero | Your personal research assistant," n.d.). A good starting point on PyTorch with Visual Recognition represent several books (Ayyadevara and Reddy, 2020)(Papa, 2021).

References

ajinkya98, 2021. Implementing CNN in PyTorch with Custom Dataset and Transfer Learning.

Ayyadevara, V.K., Reddy, Y., 2020. Modern Computer Vision with PyTorch.

Brito da Silva, L.E., Elnabarawy, I., Wunsch, D.C., 2019. A survey of adaptive resonance theory neural network models for engineering applications. Neural Netw., special Issue in Honor of the 80th Birthday of Stephen Grossberg 120, 167–203. https://doi.org/10.1016/j.neunet.2019.09.012

Brna, A.P., Brown, R.C., Connolly, P.M., Simons, S.B., Shimizu, R.E., Aguilar-Simon, M., 2019. Uncertainty-based modulation for lifelong learning. Neural Netw., special Issue in Honor of the 80th Birthday of Stephen Grossberg 120, 129–142. https://doi.org/10.1016/j.neunet.2019.09.011

Chen, Z., Liu, B., 2016. Lifelong Machine Learning. Synth. Lect. Artif. Intell. Mach. Learn. 10, 1–145. https://doi.org/10.2200/S00737ED1V01Y201610AIM033

Conda-Pack, 2021. . conda.

Conda-Pack — conda-pack 0.6.0 documentation [WWW Document], n.d. URL https://conda.github.io/conda-pack/ (accessed 12.24.21).

Containerize a conda environment in a Singularity container [WWW Document], n.d. . NewbeDEV. URL https://newbedev.com (accessed 12.20.21).

French, R.M., 1999. Catastrophic forgetting in connectionist networks. Trends Cogn. Sci. 3, 128–135. https://doi.org/10.1016/S1364-6613(99)01294-2

Furlanello, T., Lipton, Z., Tschannen, M., Itti, L., Anandkumar, A., 2018. Born Again Neural Networks, in: Proceedings of the 35th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 1607–1616.

Géron, A., 2019. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition [Book].

Grossberg, S., 2013. Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. Neural Netw., Twenty-fifth Anniversay Commemorative Issue 37, 1–47. https://doi.org/10.1016/j.neunet.2012.09.017

Grossberg, S., 1967. Nonlinear difference-differential equations in prediction and learning theory. Proc. Natl. Acad. Sci. U. S. A. 58, 1329–1334.

Grossberg, S.T., 2012. Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control. Springer Science & Business Media.

Hellström, H., Silva Jr, J.M.B. da, Fodor, V., Fischione, C., 2020. Wireless for Machine Learning. ArXiv200813492 Cs Eess.

Henning, C., 2021. Continual Learning with Hypernetworks.

hpc - Containerize a conda environment in a Singularity container [WWW Document], n.d. . Stack Overflow. URL https://stackoverflow.com/questions/54678805/containerize-a-conda-environment-in-a-singularity-container (accessed 12.24.21).
Kaushik, P., Gain, A., Kortylewski, A., Yuille, A., 2021. Understanding Catastrophic Forgetting and Remembering in Continual Learning with Optimal Relevance Mapping. ArXiv210211343 Cs.
Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., Hadsell, R., 2017. Overcoming catastrophic forgetting in neural networks. Proc. Natl. Acad. Sci. 114, 3521–3526. https://doi.org/10.1073/pnas.1611835114
Lee, S.-W., Kim, J.-H., Jun, J., Ha, J.-W., Zhang, B.-T., 2018. Overcoming Catastrophic Forgetting by Incremental Moment Matching. ArXiv170308475 Cs.
Lesort, T., Lomonaco, V., Stoian, A., Maltoni, D., Filliat, D., Díaz-Rodríguez, N., 2019. Continual Learning for Robotics: Definition, Framework, Learning Strategies, Opportunities and Challenges. ArXiv190700182 Cs.
Li, Z., Hoiem, D., 2017. Learning without Forgetting. ArXiv160609282 Cs Stat.
Mai, Z., Li, R., Jeong, J., Quispe, D., Kim, H., Sanner, S., 2021. Online Continual Learning in Image Classification: An Empirical Survey. ArXiv210110423 Cs.
Mermillod, M., Bugaiska, A., BONIN, P., 2013. The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects. Front. Psychol. 4, 504. https://doi.org/10.3389/fpsyg.2013.00504
Mundt, M., Hong, Y.W., Pliushch, I., Ramesh, V., 2020. A Wholistic View of Continual Learning with Deep Neural Networks: Forgotten Lessons and the Bridge to Active and Open World Learning. ArXiv200901797 Cs Stat.
ngailapdi, 2021. ngailapdi/LWF.
Nguyen, G., Chen, S., Do, T., Jun, T.J., Choi, H.-J., Kim, D., 2020. Dissecting Catastrophic Forgetting in Continual Learning by Deep Visualization. ArXiv200101578 Cs Stat.
Overleaf, Online LaTeX Editor [WWW Document], n.d. URL https://www.overleaf.com (accessed 12.24.21).
Papa, J., 2021. PyTorch Pocket Reference [Book] [WWW Document]. URL https://www.oreilly.com/library/view/pytorch-pocket-reference/9781492089995/ (accessed 12.19.21).
Papers with Code - Learning without Forgetting [WWW Document], n.d. URL https://paperswithcode.com/paper/learning-without-forgetting (accessed 12.20.21).
Russel, S., Norvig, P., 2020. Artificial Intelligence: A Modern Approach, 4th Global ed.

Sax, A., Zhang, J., Zamir, A., Savarese, S., Malik, J., 2019. Side-Tuning: Network Adaptation via Additive Side Networks.
sidetuning [WWW Document], n.d. . Dropbox. URL https://www.dropbox.com/sh/v7qmrj4n0yqkjjp/AAD1HNfZk8_8sdBCu-WJj-Jna?dl=0 (accessed 12.20.21).
Texmaker (free cross-platform latex editor) [WWW Document], n.d. URL https://www.xm1math.net/texmaker/ (accessed 12.24.21).
TeXstudio - A LaTeX editor [WWW Document], n.d. URL https://www.texstudio.org/ (accessed 12.24.21).
von Oswald, J., Henning, C., Sacramento, J., Grewe, B.F., 2020. Continual learning with hypernetworks. ArXiv190600695 Cs Stat.
Zotero | Your personal research assistant [WWW Document], n.d. URL https://www.zotero.org/ (accessed 12.24.21).