Thesis Overview:

Invariance and Same-Equivariance Measures for Convolutional Neural Networks

Facundo Manuel Quiroga
III-LIDI, Universidad Nacional de La Plata, Argentina
PhD in Computer Science
Advisor: Dr. Laura Cristina Lanzarini
{fquiroga,laural}@lidi.info.unlp.edu.ar

Motivation

Neural Networks are machine learning models which currently have the best performance in a wide variety of problems. Paired with optimization algorithms based on gradient descent, they can optimize thousands or millions of parameters with respect to a custom error function. Neural Networks distinguish themselves from other models but not requiring manual design of features to work correctly; instead, features are learned automatically through the optimization process, which is also called training. The design of Neural Networks is organized in layers which determine its architecture. Recently, Neural Networks with a large number of layers have been successfully trained using a set of techniques usually known as Deep Learning.

In particular, Convolutional Neural Networks, that is, Neural Networks that employ Convolutional Layers, are the state-of-the-art in most computer vision problems, including image classification. Many of the problems for which Convolutional Networks are the state-of-the-art require models to behave in a certain way whenever their input is transformed. There is a fundamental property that captures such requirement: equivariance. An equivariant function allows the prediction of the changes in its output with respect to transformations of its input.

Two special cases of equivariance are of interest:invariance and same-equivariance. Invariance states that the output of the model is not affected at all by the transformations. Same-equivariance states that the output of the model is transformed in the same way as its input.

While traditional models based on Convolutional Networks are same-equivariant to translation by design, they are neither equivariant to other transformations nor invariant to any set of transformations, at least in the usual training and testing scenarios. There are two main options to grant invariance or equivariance to a neural network model. The most traditional option has been to modify the design of the model so that it possesses such properties. The other option is to train it using data augmentation with the same set of transformations to which the equivariance is desired. However, the mechanisms by which these models acquire these properties is unclear, weather for modified models or for data augmented ones. Furthermore, the impact of the modification of the models on their efficiency and representational power is unknown. Even more, for traditional models trained with data augmentation, the different augmentation strategies to acquire equivariances have not been studied.

Objectives

Our main objective in this thesis is to contribute to the understanding and improvement of equivariance in neural network models. In terms of applications, we focus on handshape classification for sign language and other types of gestures using convolutional networks. Therefore, we set the following specific goals:

- Analyze CNN models design specifically for equivariance
- Compare specific models and data augmentation as means to obtain equivariance. Evaluate transfer learning strategies to obtain equivariant models starting with non-equivariant ones.
- Develop equivariance measures for activations or inner representations in Neural Networks. Implement those measures in an open source library. Analyze the measures behavior, and compare with existing measures.

- Characterize CNN models for image classification in terms of equivariance using the proposed measures. Compare CNN models, with or without equivariance, for handshape classification.

Given the existence of multiple methods to achieve equivariance, and the lack of deep and rigorous comparisons among them, the scope of this thesis is limited to the analysis of the models and we therefore do not propose new equivariant network models.

Contributions

The main contributions of this thesis include:

- A comparative analysis of Neural Network based models for sign language handshape classification.
- An analysis of strategies to achieve equivariance to rotations in neural networks for:
  - Comparing the performance of strategies based on data augmentation and specially designed networks and layers.
  - Determining strategies to retrain networks so that they acquire equivariance to rotations.
- A set of measures to empirically analyze the equivariance of Neural Networks, as well as any other model based on latent representations, and the corresponding:
  - Validation of the measures to establish if they are indeed measuring the purported quantity. Analysis of the different variants of the proposed measures. Analysis of the properties of the measures, in terms of their variability to transformations, models and weight initialization.
  - Analysis of the impact of several hyperparameters of the models on the structure of their equivariance, including Max Pooling layers, Batch Normalization, and kernel size. Analysis of the structure of the equivariance in several well known CNN models such as ResNet, All Convolutional and VGG. Analysis of the impact on the equivariance of using specialized models to obtain equivariance such as Transformational Invariance Pooling.
  - Analysis of the class dependency of equivariance. Analysis of the effect of varying the complexity and diversity of the transformations on the measures.

Conclusions

The main conclusion of this thesis is that the encoding of equivariance in a network is a complex process that can be understood from different perspectives, and the equivariant measures we proposed are excellent tools to investigate them.

For example, using the measures we show that the structure of the invariance of a network is a property that does not depend on the exact weights of the network, both for networks with random weights and for trained networks. However, this structure is dependent on the dataset and transformation set. On the other hand, same-equivariance is a property that does not depend on the weights nor the dataset used to measure it, and has only a mild dependence on the transformation set. This indicates that same-equivariance is determined by the network design to a larger degree than invariance. We also determine that features such as Batch Normalization layers do not affect the equivariance of the network, while the size of the kernel in Convolutional Layers does, regardless of the fact that networks with either of these characteristics achieve similar levels of accuracy.

We also propose a set of retraining experiments which can show that invariance can be encoded in different sets of layers with no effect on accuracy. This technique can also be employed to evaluate the impact of invariance in each layer separately.

Therefore, the set of measures and experimental methods we propose in this thesis allow a deep analysis of the encoding of equivariance in a network. We believe it is possible to learn more about Convolutional and Neural Networks by studying their equivariances and therefore improve existing models, enabling new applications of Computer Vision and Machine Learning.

Publications

The following publications are directly related to the contributions of this thesis:

- Facundo Quiroga y col. (2017). «A study of convolutional architectures for handshape recognition applied to sign language». En: XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017). Springer International Publishing
- Facundo Quiroga y col. (2018). «Revisiting Data Augmentation for Rotational Invariance in Convolutional Neural Networks». En: International Conference on Modelling and Simulation in Management Sciences. Springer, págs. 127-141
- Facundo Quiroga y col. (2019). «Measuring (in) variances in Convolutional Networks». En: Conference on Cloud Computing and Big Data. Springer, págs. 98-109

Additionally, during the development of this thesis we created the following resources:

- **Transformational Measures**: An open source library to compute transformational measures such as those defined in this thesis for neural network model, using the frameworks PyTorch and Numpy (https://github.com/facundoq/transformational_measures).
- **Handshape Datasets**: An open source library to facilitate the download and pre-processing of handshape datasets (https://github.com/midusi/handshape_datasets).
- **LSA16**: A database of handshape images for the Argentinian Sign Language, publicly and freely available under a creative commons license (http://facundoq.github.io/unlp/lsa16/).
- **LSA64**: A database of videos of signs of the Argentinian Sign Language, publicly and freely available under a creative commons license (http://facundoq.github.io/unlp/lsa64/)

Citation: F. Quiroga. *Thesis Overview: Invariance and Same-Equivariance Measures for Convolutional Neural Networks.* Journal of Computer Science & Technology, vol.20, no. 1, pp. 53-55, 2020.
DOI: 10.24215/16666038.20.e06.