Trainable Spectrally Initializable Matrix Transformations in Convolutional Neural Networks

*Michele Alberti[1,5], *Angela Botros[2], *Narayan Schütz[2],
Rolf Ingold[1], Marcus Liwicki[3], Mathias Seuret[1,4]

[1] Document Image and Voice Analysis Group (DIVA), University of Fribourg, Switzerland
[2] ARTORG Center for Biomedical Engineering Research, University of Bern, Switzerland
[3] EISLAB Machine Learning Luleå, University of Technology, Sweden
[4] Pattern Recognition Lab, Friedrich-Alexander-Universität at Erlangen-Nürnberg, Germany
[5] V7 Ltd, London, United Kingdom

Introduction

Vanilla CNN struggle at applying global transformations

e.g., translation, rotation, scaling, mirroring and shearing

Global transformations are very common in visual computing and signal processing, and are applied manually as a pre-processing step

e.g., Discrete Fourier Trans-formation (DFT) or Discrete Cosine Transformation (DCT)

This is surprising, considering the existing large body of literature leveraging these transformations in image processing

Motivation

Successful **CNN are in fact biased towards the texture** of the objects rather than their global shape

This suggests that the **global texture** (high frequencies) **could be more important** than the global structure (low frequencies)

Previous work in this field concluded that for this setting (**texture analysis**), we need to redesign neural architectures and devise new learning algorithms

Contribution

We provide a **proof-of-concept** and implementation for a novel architectural component, which leverages trainable linear matrix transformation module → can perform global transformations

- Open-source
- Can be initialized with spectral transformations (DCT, DFT)
- Differentiable

Our PyTorch based open-source implementations are freely available as a pip installable python package[1] and have already been integrated into the DeepDIVA[2] deep learning framework thus enabling **full reproducibility of experiments**

1. https://github.com/NarayanSchuetz/SpectralLayersPyTorch
2. https://github.com/NarayanSchuetz/DeepDIVA

Matrix Transforms in Neural Networks

We recall that a traditional neural layer has the form $\boxed{y_{nn} = f(W \cdot x + b)}$

Instead, we propose to formulate the matrix transform layer as:

$$\boxed{y_{mt} = W_1 \cdot x \cdot W_2^T}$$

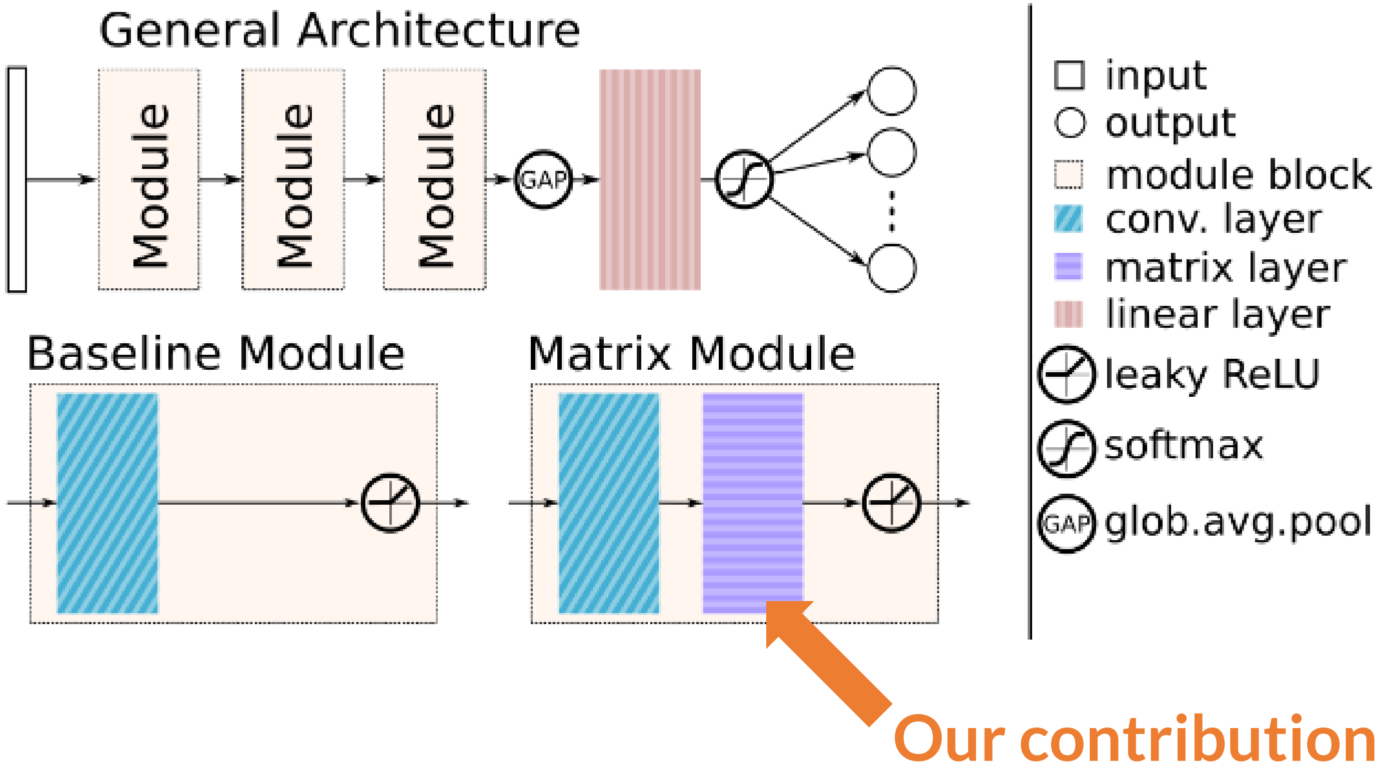
where

y_{mt}	output of the matrix transform layer $y \in Y \subseteq \mathbb{R}^{k \times l}$
x	input sample $x \in X \subseteq \mathbb{R}^{n \times m}$
W_1	left-hand weight matrix $W_1 \in \mathbb{R}^{k \times n}$
W_2	right-hand weight matrix $W_2 \in \mathbb{R}^{l \times m}$

Beyond this general form, we are we are particularly interested into the two specific settings where the matrices W_1 and W_2 are initialised s.t. they compute the2D-DFT and 2D-DCT transforms

Network Design Overview

We made experiments with several network configurations



Results Overview

Models with matrix transformations **outperform the baseline**

Matrix transforms are **beneficial for convergence**

The DFT variants appears to be the **best performing**

Conclusion and Outlook

Our component **overcomes traditional CNN limitation** and enables applying global transformations

Empirical results shows this is **beneficial in terms of learning speed and final performances** on a image classification task

Spectral initialization as DCT or DFT brings substantial speedups in terms of convergence, when compared to random initialization

Our experiments are a proof-of-concept and are limited to small networks/datasets

(partly due to a heavy hyper-parameter optimization setup)