Convolutional Neural Networks from Image Markers

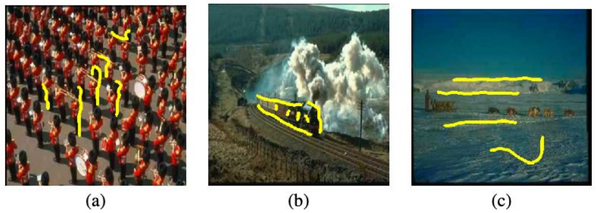
Bárbara C. Benato, Italos S. Estilon, Felipe L. Galvão, Alexandre X. Falcão
{barbara.benato, italos.estilon, felipe.galvao, afalcao}@ic.unicamp.br
University of Campinas

Abstract

A technique named *Feature Learning from Image Markers* (FLIM) was recently proposed to estimate convolutional filters, with no backpropagation, from strokes drawn by a user on very few images (e.g., 1-3) per class. This paper extends FLIM for fully connected layers and demonstrates it on different image classification problems. The work evaluates marker selection from multiple users and the impact of adding a fully connected layer. The results show that FLIM-based convolutional neural networks can outperform the same architecture trained from scratch by backpropagation.

Motivation

Convolutional neural networks (CNNs) have shown remarkable performance in image classification problems. However, CNNs may present complex and deep architectures, requiring considerable human effort in data annotation, and resulting in non-explainable models. Recently, user involvement seems crucial to discover more efficient and effective ways to transfer human knowledge to machines during the deep learning process. One example, FLIM, can estimate relevant filters to compose a given number of convolutional layers from strokes drawn by a user on image. **The strokes are drawn on image regions that best represent the classes, in a way that the user guides the network training.**



(a) (b) (c)

Figure 1: Marker selection (yellow) on images of three out of six classes, as defined for this work from a Corel Stock CD with JPEG images: (a) royal guard, (b) train, and (c) snow.

Objectives

Design of Convolutional Neural Networks with:

- user control and understanding about the learning process;
- reduced user effort in data annotation; and
- the required number of convolutional layers.

In this work, we present a methodology consistent with the above aims and extend it to incorporate fully connected layers.

Method: Feature Learning from Image Markers

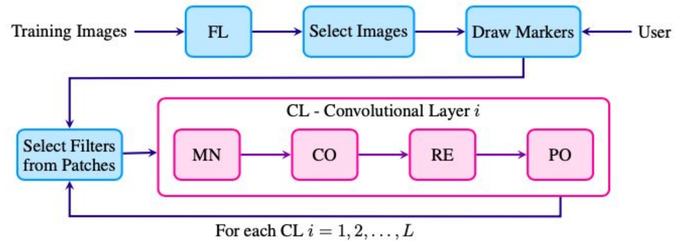
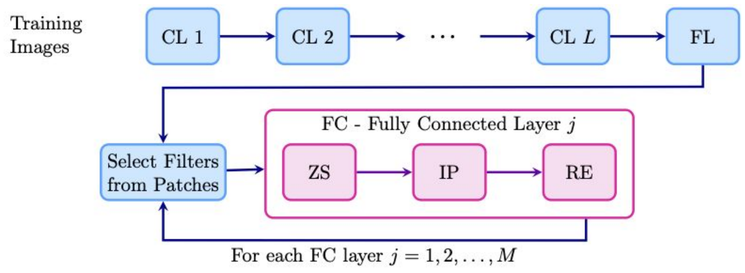


Figure 2: Building one convolutional layer after another with FLIM. FL stands for flattening, MN for *marker-based normalization*, CO for convolution, RE for ReLU activation, and PO for max pooling.

CNNs from Image Markers



Proposed CNN from Image Markers, in which CL stands for convolutional layer, FL for flattening, ZS for z-score normalization, IP for inner product, and RE for ReLU.

Experimental set-up

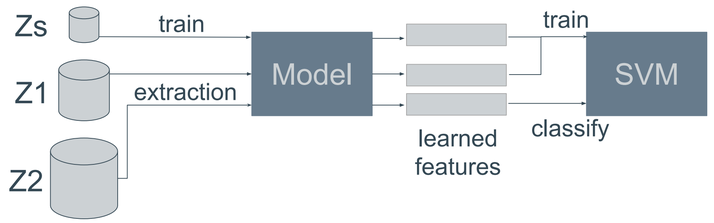
We used k-means for clustering and markers drawn by two users. We selected three datasets to evaluate FLIM: Citrus Leaves (604 images), CorelStock CD (355 images, a subset), Rock, Paper, and Scissors (2892 images). All images have been rescaled to 400x400 pixels. Each dataset Z was randomly and stratified split into sets Z1 and Z2, three times:

Z1: 30% of training samples -> **Zs:** very small subset to draw markers

Z2: 60% of testing samples

dataset	classes	$\lvert Z_s \rvert$	% of $\lvert Z \rvert$	$\lvert Z_1 \rvert$	$\lvert Z_2 \rvert$	$\lvert Z \rvert$
Corel	6	13	3.70%	104	251	355
Citrus	5	9	1.50%	179	425	604
RPS	3	6	0.02%	867	2025	2892

A drawing tool -- such as a free-hand brush -- was used by each user (A and B) to draw strokes onf Zs. we compare our model FLIM with a baseline model. This baseline model has the same FLIM's architecture and was trained by backpropagation with linear learning rate decay.



Experiments and Results

database		split1 $CL1$	split1 $CL2$	split1 FC	split2 $CL1$	split2 $CL2$	split2 FC	split3 $CL1$	split3 $CL2$	split3 FC	mean $CL1$	mean $CL2$	mean FC
Corel	A	0.9084	**0.9203**	0.9004	0.9363	**0.9442**	0.9203	0.9004	0.8964	0.8964	0.9150	**0.9203**	0.9057
	B	0.8884	0.9043	0.8844	0.9322	0.9362	0.9083	0.9043	0.9163	**0.9243**	0.9083	0.9189	0.9057
	bp	-	-	0.8690	-	-	0.8460	-	-	0.8170	-	-	0.8438
Citrus	A	0.7859	0.8118	**0.8400**	0.7671	0.7718	**0.8306**	0.7859	0.7788	0.8024	0.7796	0.7875	0.8243
	B	0.7317	0.7764	0.8282	0.7858	0.7976	0.8258	0.8000	0.8305	**0.8352**	0.7725	0.8015	**0.8298**
	bp	-	-	0.7110	-	-	0.6870	-	-	0.7090	-	-	0.7022
RPS	A	**0.9891**	-	0.9877	0.9852	-	**0.9896**	0.9921	-	**0.9965**	0.9888	-	**0.9913**
	B	0.9827	-	0.9866	0.9767	-	**0.9896**	0.9906	-	0.9911	0.9833	-	0.9891
	bp	-	-	0.9750	-	-	0.9790	-	-	0.9770	-	-	0.9771

- FLIM-based CNN (with and without FC layer) outperformed the CNN trained by backpropagation (*bp*) in all datasets, splits and on average;
- The strokes on representative parts of the classes seem to be the best strategy, as followed by user A, but the additional selection of markers on parts that do not represent classes, as followed by user B, was better in some splits.

Conclusion

The experiments have demonstrated that: (i) **it is possible to effectively train CNNs from image markers on representative regions of the classes with no backpropagation**, (ii) FLIM-based CNNs can outperform a same CNN architecture trained by backpropagation, and (iii) a fully connected layer may improve the results of the FLIM-based CNN with convolutional layers only.

Not only the number of selected images was small, but also the number of markers per image was small (e.g., less than 10). **As pros**, FLIM allows an intuitive, straightforward, and explainable mechanism to train CNNs with reduced user effort. **As cons**, for problems with many classes, marker selection in all classes requires more user effort.

Acknowledgments

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)* - Finance Code 001; by *Petróleo Brasileiro S.A. (PETROBRAS)* and *Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP)* grants #4600556376 and #4600583791; and by *FAPESP* grants #2014/12236-1, and #2019/10705-8, *CNPq* grants 303808/2018-7. The views expressed are those of the authors and do not reflect the official policy or position of the *São Paulo Research Foundation*.