Knowledge Transfer for Melanoma Screening with Deep Learning (Reimplementation by Swathi Guptha)

Afonso Menegola, Michel Fornaciali, Ramon Pires,
Flavia Vasques Bittencourt, Sandra Avila, Eduardo Valle

RECOD Lab, DCA, FEEC, University of Campinas (Unicamp), Brazil
RECOD Lab, IC, University of Campinas (Unicamp), Brazil
School of Medicine, Federal University of Minas Gerais (UFMG), Brazil

Abstract

Deep learning's greed for large amounts of training data poses a challenge for medical tasks, which we can alleviate by recycling knowledge from models trained on different tasks,in a scheme called transfer learning. This paper deals with analysing the impact of transfer learning while training a model using a pre-trained model. In order to analyse the impact of transfer learning various combinations of source dataset and Target dataset is used. Reimplementation code can be found at: Github

1. Introduction

From all skin cancers, Melanoma prognosis just represents just 1% of cases, but 75% of deaths. Melanoma's prognosis is good when detected early, but deteriorates fast as the disease progresses. The disease's rarity, lethality, fast prognosis and diagnosis subtlety make the automated screening challenging. Various techniques like computer vision techniques which deals with extracting low level local features and feeding to a classifier[1],training deep neural network (DNN) from scratch,segmenting lesions before feature extraction is also common choice for melanoma screening [2,3] which is eschewed while using DNN [1,2],deep neural networks plus transfer learning [2,3] were used in order to address the automated screening of melanoma. The main contribution of this paper is studying it in more detail.

1.1. Dataset

The datasets employed to train and test the target models (melanoma screening) were the Interactive Atlas of Dermoscopy (Atlas), and the ISBI Challenge 2016.The sources datasets employed for the transfer (pre-training of the DNNs) were the Kaggle Challenge for Diabetic Retinopathy Detection dataset, with a training set of 35,000+ high-resolution retina images ; and the ImageNet dataset containing 1M training images labeled into 1,000 categories. The use of ImageNet dataset was indirect because training over it is so time consuming, the authors opted for employing pre-trained source DNNs.

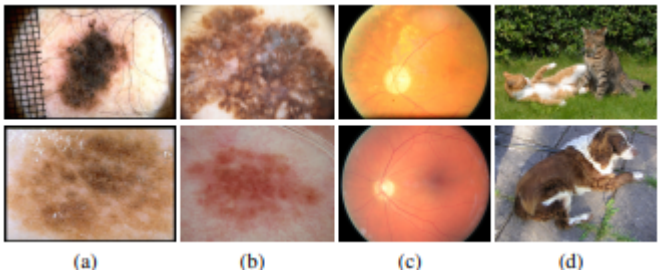

Figure 1: Samples from datasets used here: (a) Atlas; (b) ISIC; (c)Retinopathy; (d) ImageNet. In this paper,datasets c and d are source datasets used for transferring knowledge to target models trained in the target task of melanoma screening,trained and evaluated in datasets a and b.

1.2. Methodology

The method used by the authors in detail is:
1. To reach the input size required by VGG,all the images are resized to 224 × 224 pixels
2. The inputs are "re-centered", as required by VGG, by subtracting the mean of the training dataset.
3. Training set is balanced by augmenting the data for the minority classes, applying a random transformation (scaling, rotating, or flipping the images).The random number generator is fixed at the very beginning of each experiment, so that the training sequence is the same for all experiments. The model is trained for 5X2 folds with 60 epochs, learning rate 10^{-3} , Nesterov momentum 0.1,

VGG-M	Datasets	
Model	**Source Dataset**	**Target Dataset**
A	Trainig Atlas from scratch (No transfer lesrning)	
B	Retinopathy	Atlas (without FT)
C	Retinopathy	Atals (with FT)
D	Imagenet	Atlas (without FT)
E	Imagenet	Atlas (with FT)
F	Imagenet+Retinopathy	Atlas (with FT)

Figure 2: Models Trained (FT = fine tuning)

Experimental Design	No Transfer	From Retinopathy		From ImageNet		Double Transfer
		no FT	with FT	no FT	with FT	with FT
Malignant *vs.* Benign	76.0	72.8	76.0	79.1	**82.5**	78.8
Melanoma *vs.* Benign	75.7	73.5	75.3	77.9	**80.9**	**80.9**
Melanoma *vs.* Carcinoma *vs.* Benign	73.0	71.4	72.8	79.4	**83.6**	81.8

Figure 3: Results from the paper

Experimental Design	No Transfer	From Imagenet	
		No FT	FT
Malignant vs. Benign	62.3	73.2	**80.2**
Melanoma vs. Benign	60.7	76.1	**79.0**
Melanoma vs. Carcinoma vs. Benign	59.3	75.3	**82.5**

Figure 4: Results by re-implementation

and L2 regularization of 5×10^{-3}. A 10% stratified split of the training set is performed for validation.The model minimizes validation loss
4. The Layer 19 output is extracted, which are vectors of 4, 096 dimensions, and use that output as a feature-vector to describe the images; and feed them to a linear SVM classifier (Sklearn implementation) to make the decision on the skin lesion.
Few skin cancers, notably basal cell carcinomas become a challenge for that model: should they just be put in the negative class with all other lesions.Three experimental designs were evaluated to address the challenge, varying the labeling of the classes:
• Malignant vs. Benign lesions: melanomas and basal cell carcinomas were considered positive cases and all other diagnoses were negative cases;
• Melanoma vs. Benign lesions: melanomas were positive cases while all other diagnoses were negative ones, removing basal cell carcinomas;
• Basal cell carcinoma vs. Melanoma vs. Benign lesions: here we have three classes, with all other diagnoses under a single Benign label

1.3. Issues Faced

The major issue faced while re-implementing the github repo is the usage of old package Theano, Lasagne and nolearn. The modules needed special configuration to use GPU and the method of configuration became obsolete now. In order to complete the task I had to redo the entire work using keras. Since the retinopathy dataset was 80GB large so it wasn't feasible for me to implement model B,C,F.

1.4. Results

The main metric was the Area Under the ROC Curve (AUC); for the design with three classes,three one-vs-one AUCs are computed and reported their average.The figure 3 representing the research paper results shows that transfer learning from Retinopathy leads to worse results than transferring from the general task of ImageNet, even in combination with the latter. That might indicate that transferring from very specific tasks poses special challenges for overcoming the specialization. The best protocol we found was to simply transfer from ImageNet, with finetuning.

In fig 4, it's evident that Imagenet with fine tuning gives best results when compared to all the re-implemented results. The slight difference between the result is because few images in the dataset were missing. In the model trained from scratch, the vast difference between the original and re-implemented results is mainly because the model wasn't fully trained because it was throwing out of memory error on ada after 150 epochs which I couldn't resolve.

1.5. Conclusion

The paper expected that transfer learning from a related task (in this case, from Retinopathy]) would lead to better results, especially in the double transfer scheme, that had access to all information from ImageNet as well. However, the results showed the opposite. (suggesting that adaptation from very specific — even if related — tasks poses specific challenges. Still, the authors believe that further investigation is needed).

1.6. Final Thoughts

The approach used in paper isn't upto the mark when compared to present methods. The paper extracts features from a CNN and gives those features to SVM for classification. More finer results could be obtained if a pre-trained VGG19 or VGG16 model is used.

1.7. References

[1] M. Fornaciali, M. Carvalho, F. Vasques Bittencourt, S. Avila, and E. Valle, "Towards automated melanoma screening: Proper computer vision reliable results," arXiv preprint arXiv:1604.04024, 2016.

[2] X. Sun, J Yang, M. Sun, and K. Wang, "A benchmark for automatic visual classification of clinical skin disease images," in ECCV, 2016, pp. 206–222.

[3] N. Codella, Q.-B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. Halpern, and J. Smith, "Deep learning ensembles for melanoma recognition in dermoscopy images," IBM J Res Dev, vol. 61, 2017.