A Report on Orientation Estimation

groveh

March 2019

1 Introduction

Straightening photos to their upright position is a fundamental problem that has numerous practical outlets, from allowing the interpretation of documents to correcting vacation photos. The priority of being able to read scanned documents lead to a solution that was found long ago. These methods, however, exploit the unique structure of document images such as the layout of each line or the shapes of letters. Other areas include simplified angle estimation, such as inferring the optimal orientation in 90 degree increments, or performing image transforms and using lines to correct rotation. This paper presents various networks to infer the exact angle of rotated images.

2 Data

Data augmentation is crucial to preserve the generality of the features learned by the network. While a ground truth is known from annotation, this is used to orient the picture at a true zero degree rotation at its natural orientation. From here artificial rotation transformations can be made with the intent to learn these characteristics. This is more beneficial than using the annotated angle because at each epoch we can perform a random transformation rather than the same one each run through the data. This reduces overfitting and increases the generality of the features learned by the network so that it is well calibrated for unseen data.

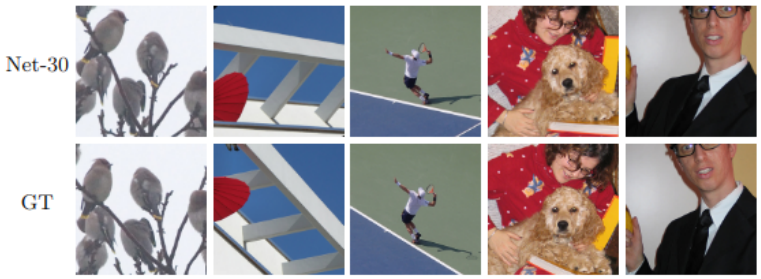
3 Networks

This paper considered the orientation problem at three different difficulty levels. The easiest and middle setting expect the input image to be at most $\pm 30^\circ$ and $\pm 45^\circ$ respectively from ground truth orientation. The middle stage is inferred to be more difficult because it introduces possible confusion between horizontal and vertical lines. The most difficult setting estimates all angles in the range of $[0, 360)^\circ$, therefore it has no knowledge about course orientation, such as landscape or portrait. All networks were built using AlexNet architecture and

pre-trained on ImageNet. In summary, AlexNet consists of 5 convolutional layers followed by 3 fully connected layers. A ReLU activation is applied after each fully connected layer, normalization and dropout are applied, and L1 loss is used. Pre-training from ImageNet also improves results because the class labels help learn semantic features that are too complex to learn from the orientation objective alone. The output layer was changed to have two output units, one for both positive and negative rotations. Therefore the output vector is represented as $[max(0, \alpha), max(0, -\alpha)]$. In addition, a classification network with 4 outputs corresponding to 0, 90, 180, and 270 degrees was implemented with softmax output activation and cross-correlation loss. It is used for course estimation before estimating using the 3 regression networks described.

4 Results

Neural network orientation correction uses abstract features that contradicts a humans intuitive inference. True image orientation estimation cannot rely on any image features being present if it allows input of any natural image. Deep convolutional networks learn their own general features across training images, sometimes including various lines and objects that humans would know how to orient based off of previous knowledge. Below are examples where lines from within the image lead to a skewed result:



While these examples may be considered similar enough to accept, it shows that algorithms trained with specific features in mind cannot be expected to generalize the orientation estimation problem well enough to be reliable on all natural images. Other than small deviations, overall results are as follows:

Task	Net-30	Net-45	Net-360	Net-rough+45	Hough-var	Hough-pow	Fourier
±30°-all	**3.00**	4.00	19.74	19.64	11.41	10.62	10.66
±30°-easy	**2.17**	2.83	19.48	17.90	8.44	7.04	8.64
±30°-hard	**4.26**	5.75	20.12	22.24	15.88	15.99	13.69
±45°-all	-	**4.63**	20.64	19.24	16.92	13.06	16.51
±45°-easy	-	**3.24**	21.26	19.29	14.08	9.01	13.32
±45°-hard	-	**6.71**	19.70	19.15	21.16	19.13	21.28
±180°-all	-	-	**20.97**	**18.68**	-	-	-
±180°-easy	-	-	**20.29**	**18.70**	-	-	-
±180°-hard	-	-	**21.98**	**18.65**	-	-	-

With the -all, -easy, and -hard tags represent easy images to orient, hard images to orient, and the combined set of images. Other baseline methods were also tested for comparison, such as two variations of a Hough transform, and a Fourier transform. The two easiest network settings provide an acceptable result, roughly 2-3 degrees difference from ground truth. While it is not a perfect outcome, most well designed detection networks will work near optimally with this slight angle deviation.