CNN Assistance in Jigsaw Puzzle Solution

Tianbao Li
Department of Computer Science
University of Toronto
Toronto, ON M5S 1A1
tianbao@cs.toronto.edu

Abstract

This paper presents an innovative way to assist solving jigsaw puzzle with the help of deep convolutional neural networks. The main goal of this project is to predict whether two pieces from the jigsaw puzzle should be neighbors in the origin image. Proceeding from traditional methods of using color (like RGB) distance, here, we introduce the low level features in the image which are extracted by convolutional neural networks. Compared with color-based solutions, using the feature maps generated can help with a deeper intuition on the correlation between edges. The proposed algorithm can achieve considerable accuracy on adjacency prediction.

1 Introduction

Jigsaw puzzles were first introduced around 1760 for map research and then became a popular intelligence entertainment [4] in the last few centuries. The origin image is divided into $N \times M$. To solve this, people need to cluster similar tiles, find the neighbors and then reconstruct the origin image. However, due to the possible locations and relationship behind pieces, there are quite a huge number of solutions ever for a small jigsaw puzzle, for example, $(8 * 8)! = 1.2688693 * 10^{89}$ solutions for an $8 * 8$ puzzle. Indeed, this problem has be proven to be a NP-complete one [1, 2]. Though tough, actually, jigsaw puzzle solving is quite meaningful beyond the intelligence challenge. It helps a lot in combining shredded of documents [9, 10] and even recovering artifacts debris[7].

Recently year, there have been a lot researchers working on this problem and achieved much progress. However, one the the most important sub-problems in this procedure, determining adjacency between pieces, seems keeps in the same track for long. When people solve the puzzle manually, they need needs image information at edges, such as color, texture, instance, etc. to make the judgment whether two pieces are neighbors. However, most of previous research only eyes on the color features such as RGB distance. This make such solutions quite bad for simple-colored images, especially in reconstructing printed documents of ancient artifacts.

Here, we want to mimic how people really solve jigsaw puzzles and work on finding features capable to help make adjacent prediction. Recently, Convolution Neural Networks (CNN) leads the research area of computer vision with its capability of extracting inside features underneath the images and sole hard problems such as detection[13] and segmentation[5, 18]. Based on the existing tremendous structures, people start to apply transfer learning and use extracted features to solve related computer vision problems[12].

In this paper, we contribute in providing neural networks to predict whether two tiles from a jigsaw puzzle are adjacent in the origin image. The neural networks judge the jigsaw piece pair by extracted low lever image features from pre-trained CNN together with color information at edges, and outputs whether this two tiles are neighbors. This model achieves outstanding prediction accuracy in images from ILSVRC2012 dataset[14].

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

2 Related Work

Jigsaw has been researched on for many years. One most basic idea is to evaluate the compatibility of the adjacent pieces and take a strategy, such as greedy search, to arrange the pieces. One famous work is the Genetic Algorithm (GA) [15]. Given initial candidate solutions, it applies operations like selection, reproduction and mutation based on the color-distance fitness function. Similarly but innovative, [16] first introduces deep neural network to jigsaw solver and transforms the puzzle to the piece pair adjacency prediction. It samples piece edges to learn the adjacent likelihood through DNN based on the color distance. For these two solvers, they only use color information to judge whether two pieces should be together. However, human use some others like texture to solve. Also, though high accuracy, these solution could not solve the jigsaw puzzle which generated by single-colored or colorless images. So, there are still some further steps on it.

Recently, with the prosper of convolutional neural networks in computer vision area, convolutional neural networks (CNN) also becomes a good tool to solve jigsaw puzzles. Because of the fact that a good CNN is hard to train and the training data needs many images and well labels, more and more researchers choose the apply transfer learning on pre-trained models, fine turning on the last a few layers or using it as a feature extractor, instead of go over all the process [12]. CNN can extract regional features in many perspectives, such as color, texture, pattern, instance. Features in the formers usually reflect the basic information of the images, while the latter layers can make high-level judgement. These can be good inference for judging adjacent pieces. So, [3, 11] choose to use feature maps from pre-trained CFN [11] (siamese-ennead AlexNet [8]), VGG [6] or Resnet [17] and predict the location. However, the main problem for these approach is that they hold a siamese structure with shared weights for each location, which means they can only solve a limited number for pieces (3×3 in [11], 2×2 and 2×3 in [3]). With the piece amount increasing, the network becomes extremely heavy to train.

3 Preliminary

3.1 Problem Definition

Jigsaw puzzles aims to reshape multiple non-overlapping pieces from the origin image to the correct arrangement. In our project, like in most computer research, we focus on equal-sized squared $W = N \times M$ pieces. It refuses the aid of shape matching but asks for more on image information. The expected result is a set of indexes $I = (p_1, p_2, \ldots, p_W)$ which p_i is the corresponding index for each piece in the origin image. For the worst case, it takes the complicity of $O(W!)$. For example, Figure 1 is a $3 * 3$ jigsaw puzzle. The left subgraph is the origin image, and the right subgraph is the puzzle which $I = \{6, 2, 7, 3, 5, 4, 0, 8, 1\}$.

3.2 Motivation

Starting from the idea in [16], instead of solving the jigsaw puzzle from the from to the end, we choose to focus on the prediction whether two jigsaw puzzle tile should be neighbor in the origin image. In recent research, nearly all of the researchers decide to use color information to judge whether two tiles are adjacent. And as in the result, the color distance like RGB distance appears as a good measurement. In some work[15] the color distance is defined as the sum of squared color differences in three-dimensional color space like RGB. For image piece x_i and x_j as a $K \times K \times 3$ matrix which K is the edge length and cb is the color band, the distance is

$$D(x_i, x_j, r) = \sqrt{\sum_{k=1}^{K} \sum_{cb=1}^{3} (x_i(k, K, cb) - x_j(k, 1, cb))^2}$$

However, this color-based measurement doesn't work all the time. For example, in Figure 2, RGB-based can not predict the similarity of gray-scaled images. Also even changed it to gray-level distance, it doesn't work so well. Also, like in Figure 3, even in RGB band, main part of the image is in color blue, even all blue pixels in some edges. So the distances are so close for all edge pairs, which can leads to high error rate for the prediction. Moreover, for the images of furs, hairs and others with obvious pattern structure, like in Figure 4, it is easy to have a low distance with the displacement fitting the pattern though it is wrong. Only color information can not have a good result on this.

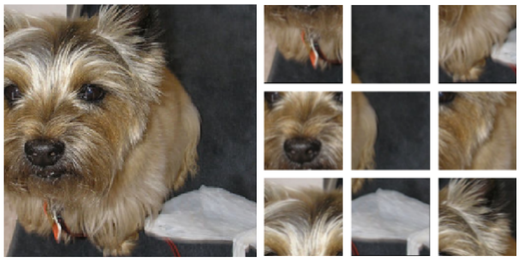

Figure 1: Colorful Jigsaw Puzzle Example

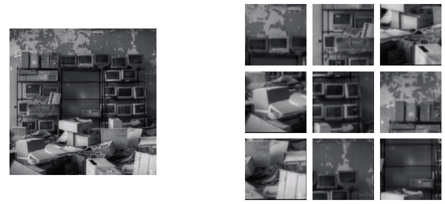

Figure 2: Black Jigsaw Puzzle Example

Figure 3: Blue Jigsaw Puzzle Example

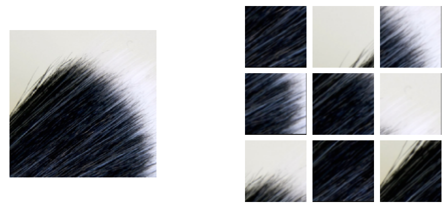

Figure 4: Textural Jigsaw Puzzle Example

Here, we combine the features extracted from Resnet34[6]. Usually, people choose the extract the feature maps after the activation functions which contains much image information such as color, texture and so on. Resnet34 is constructed on the structure of BasicBlock (several layers of convolutions, batch normalization, ReLUs, etc) and can skip some of the blocks to achieve the residual learning with more blocks. Resnet34 has 3, 4, 6 and 3 blocks in each section. There are some discussions in [11] that feature maps from low layers of CNN can help with solving the jigsaw puzzle better. In our algorithm, we choose to use the ReLU output after the first block.

Combining the measurements above, the goal of the algorithm is to build the neural networks based classifier to predict the adjacent likelihood of a pair of two jigsaw puzzle edges in the origin image. Note that not all pixels help with the prediction near the edge, we only use the feature maps and color value for the two rows near the edge to reduce the work load of the neural networks.

4 BuddyNet

In this section, we will discuss the neural network structure, BuddyNet. It is a NN classifier which predicts the adjacency likelihood of a pair of jigsaw tiles in the origin image.

5 RGBNet

6 Evaluation

6.1 Dataset

For a jigsaw puzzle with N pieces each edge,

References

[1] Tom Altman. Solving the jigsaw puzzle problem in linear time. *Applied Artificial Intelligence an International Journal*, 3(4):453–462, 1989.

[2] Erik D Demaine and Martin L Demaine. Jigsaw puzzles, edge matching, and polyomino packing: Connections and complexity. *Graphs and Combinatorics*, 23(1):195–208, 2007.

[3] Lucio Dery, Robel Mengistu, and Oluwasanya Awe. Neural combinatorial optimization for solving jigsaw puzzles: A step towards unsupervised pre-training. 20167.

[4] Herbert Freeman and L Garder. Apictorial jigsaw puzzles: The computer solution of a problem in pattern recognition. *IEEE Transactions on Electronic Computers*, (2):118–127, 1964.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2980–2988. IEEE, 2017.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] David Koller and Marc Levoy. Computer-aided reconstruction and new matches in the forma urbis romae. *Bullettino Della Commissione Archeologica Comunale di Roma*, 2:103–125, 2006.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[9] Samuel R Levin. The computer and literary studies, 1975.

[10] Marlos AO Marques and Cinthia OA Freitas. Reconstructing strip-shredded documents using color as feature matching. In *Proceedings of the 2009 ACM symposium on Applied Computing*, pages 893–894. ACM, 2009.

[11] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European Conference on Computer Vision*, pages 69–84. Springer, 2016.

[12] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 512–519. IEEE, 2014.

[13] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. *arXiv preprint*, 2017.

[14] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[15] Dror Sholomon, Omid David, and Nathan S Netanyahu. A genetic algorithm-based solver for very large jigsaw puzzles. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1767–1774. IEEE, 2013.

[16] Dror Sholomon, Omid E David, and Nathan S Netanyahu. Dnn-buddies: a deep neural network-based estimation metric for the jigsaw puzzle problem. In *International Conference on Artificial Neural Networks*, pages 170–178. Springer, 2016.

[17] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[18] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.