Bi-stream Region Ensemble Network: Promoting Accuracy in Fingertip Localization from Stereo Images

Cairong Zhang
zcr17@mails.tsinghua.edu.cn
Guijin Wang
wangguijin@tsinghua.edu.cn
Xinghao Chen
chen-xh13@mails.tsinghua.edu.cn
Huazhong Yang
yanghz@tsinghua.edu.cn

Department of Electronic Engineering
Tsinghua University
Beijing, China

Abstract

Accurate fingertip localization is an important and challenging problem in Human Computer Interaction (HCI). Previous research on hand pose estimation are mainly based on depth images. However, due to the noise and missing values in depth images, it is difficult to improve the performance of fingertip localization. In this paper, we propose a novel scheme named Bi-stream Region Ensemble Network (Bi-REN) to estimate the locations of fingertips and the wrist from stereo images directly, without converting them into depth images. It first extracts feature maps using DenseNet structure in two streams from left and right images separately, and then regions of feature maps are cropped. After corresponding feature regions of the two streams are concatenated, they are fused by fully connected layers to predict the final joint positions. Our method achieves the highest accuracy in a publicly available dataset ThuHand17, with the mean error of the six joints (the wrist and five fingertips) as 8.98mm, reduced around 18% compared with the state-of-the-art method.

1 Introduction

Hand pose estimation is the essential technique in Human Computer Interaction (HCI), for example, virtual reality, augmented reality and remote control. Recently hand pose estimation based on depth images has drawn lots of attention from researchers [5, 6, 12, 14, 15, 16, 18, 22, 23]. Fingertip localization is much more difficult than localizing other joints in a hand due to large variation of viewpoints, high flexibility of fingertips, and poor depth quality around fingertips in depth images [17]. Estimation of hand poses or fingertip locations can also be done based on stereo images instead of noisy depth images. But traditional methods estimating hand poses from stereo images [24] first convert stereo images into depth images [13, 19], and then hand poses are estimated. However, large noise may be introduced when

depth images are acquired from stereo ones, causing low quality of depth values around fingertips. Due to error accumulation in calculation of disparities and pose estimation from depth images, the accuracy of fingertip localization is hindered.

Therefore, it is beneficial to estimate hand poses or fingertip locations from stereo images directly. Several methods are used to estimate fingertip locations from binocular images directly [4, 21], without converting binocular images into depth images. In [4], hand mask images are extracted from binocular images and fed into a deep convolution neural network (CNN) to predict the 3D positions of fingertips and the palm center. But the network architecture is quite simple and cannot utilize the features of binocular images effectively. Besides, lots of information is lost with only mask images as input. In [21], original images and hand mask images are both used as the input considering that mask images are more robust to illumination and skin colors, and original images contain more information of the hand. Low level feature maps of both left and right images are extracted by convolution layers with the same structure and shared parameters, while high level features of the two streams are extracted by CNN layers with the same structure but different parameters separately. The two-stream method outperforms Chen et al. [4] on the ThuHand17 dataset [3].

In this paper, we propose a novel scheme named Bi-stream Region Ensemble Network (Bi-REN), to localize the fingertips (and the wrist) from stereo images directly. Cropped hand mask images are obtained from stereo images, and fed into the network along with cropped stereo images. Inside the network, feature maps are extracted using the structure of DenseNet [9] in two-stream style. To the best of our knowledge, we are the first to incorporate DenseNet structure into hand pose estimation. Suppose that the features of the left and right image are similar, Bi-REN extracts the similar features of left images and right images by DenseNet structured convolution layers with the same structure and shared parameters. After feature maps of the two streams are extracted, they are cropped into nine grid regions following [20]. The cropped feature regions of two streams are fused by concatenation and then forwarded into fully connected (FC) layers to estimate hand poses. Bi-REN achieves state-of-the-art performance over existing methods in fingertip localization from stereo images on the ThuHand17 dataset [21], with the mean error of the six joints reduced from 10.91mm [21] to 8.98mm (relatively 18%).

The remainder of this paper is organized as follows. In Section 2, we present details about our Bi-REN. In Section 3, we briefly introduce the techniques we used to perform preprocessing on the binocular image dataset. Experimental results and comparison among the different models are provided in Section 4. Section 5 concludes this paper and discusses the future work.

2 Bi-REN

In this section, we describe the proposed Bi-REN, which estimates the positions of fingertips and the wrist from stereo images directly without conversion into depth images.

2.1 Framework

As shown in Figure 1, Bi-REN uses cropped masks and cropped stereo images as input. First, left image and left mask are concatenated, as well as right image and right mask. Left inputs and right inputs are processed in two different streams with the same structure and shared parameters.

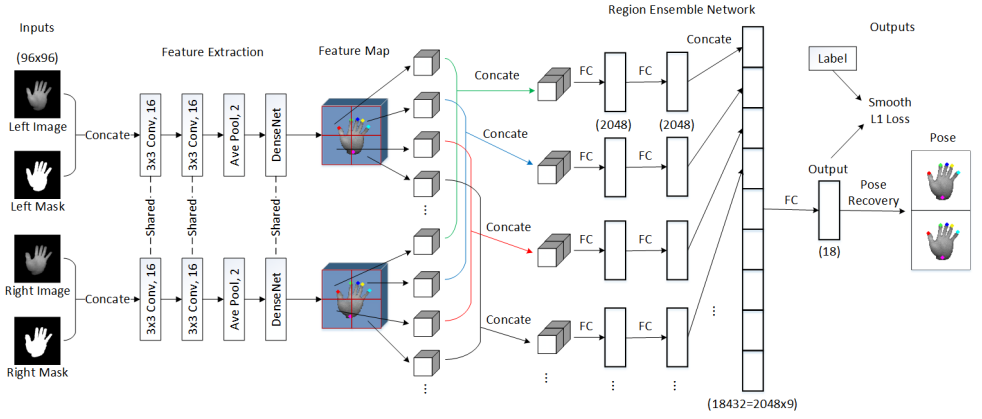

Figure 1: The framework of our Bi-stream Region Ensemble Network (Bi-REN). "$k \times k$ *Conv*$,c$" represents a convolution layer with a $k \times k$ size kernel and c channels ($stride = 1, pad = 1$). "*Ave Pool*$,2$" denotes an average pooling layer with kernel size 2×2 ($stride = 2, no\ padding$). "FC" represents a fully connected layer. The numbers inside parentheses denote the sizes of images or vectors. Only 4 regions in each stream are shown, see [20] for setting of all nine regions.

In feature extraction module, DenseNet structured layers are employed to extract feature maps of left input and right input (both in size $96 \times 96 \times 2$) in two streams separately. We stress the differences between our feature extraction module and the standard DenseNet. Before the input images forwarded into dense blocks in DenseNet, they first pass through two convolution layers and an average pooling layer. This modification eases the consumption of GPU memory since DenseNet is high memory consumed. Furthermore, we remove batch normalization (BN) [10] layers used in standard DenseNet, considering that BN helps little in regression tasks.

Inspired by REN [20], we utilize region ensemble method to localize fingertips from stereo images. Nine feature regions are extracted from the feature maps in each stream and every two corresponding regions of left and right feature maps are fused by concatenation. Then the nine concatenated regions are integrated by FC layers. Through another FC layer and an additional pose recovery layer transforming the 18-dimension output into hand poses, Bi-REN predicts hand poses end-to-end. We use rectified linear unit (ReLU) [8] as activation function.

Our Bi-REN differs from REN (for depth images) in three aspects. (1) We employ DenseNet [9] in feature extraction instead of CNN with residual connections in REN. (2) Bi-REN is a two-stream network applicable to stereo images, which extracts feature maps of left inputs and right inputs in two streams. (3) After grid regions extracted from two-stream feature maps separately, each pair of corresponding regions are concatenated to fuse the information of left and right images.

2.2 Inputs

The input of Bi-REN is four images: the stereo images and the mask images (both left and right) after preprocessing, which we will discuss later in Section 3. Besides, in order to

increase the model robustness of different sizes of hand shapes, we use multi-scale training. Three sizes of images are cropped from the original stereo images and binary mask images: 240×240, 200×200, 160×160. So the training set is enlarged by three times. While testing, the size of 200×200 is used to crop the images. The cropped left image and cropped left mask are concatenated into a two-channel input, while corresponding crops of right ones concatenated as well.

Data augmentation To further increase the robustness of Bi-REN, we perform data augmentation by translating the images for random (m,n) pixels along both horizontal and vertical directions during training (m and n are sampled from uniformed distributions with predefined borders). The ground truths of the six joints are translated correspondingly.

2.3 Two Stream DenseNet Structure for Feature Extraction

We use two streams with the same structure and shared parameters to extract feature maps of left images and right images separately. Bi-REN incorporates DenseNet structure into feature extraction. As explained in [9], all layers in DenseNet are connected directly with each other, with the feature maps of different layers combined by concatenation. But unlike [9], in Bi-REN, two convolution layers and an average pooling layer are used in front of the first dense block. The BN layers used in standard DenseNet are also removed in Bi-REN. Besides, we set a small growth rate and less layers in a dense block than [9].

Specifically, in our Bi-REN, the growth rate of a dense block is 24. There are 3 dense blocks, each containing 2 convolution layers with kernel size 3×3, stride 1 and zero-padded to keep the feature map size fixed. The first convolution layer in the first dense block has 16 output channels. Between two contiguous dense blocks, a 1×1 convolution and a 2×2 average pooling layer with stride 2 are set as the translation layers. Considering the first two convolution layers before entering dense blocks, there are totally 10 convolution layers in each stream in our Bi-REN.

2.4 Region Ensemble Network

Inspired by REN [20], which promotes the accuracy in hand pose estimation from depth images, we apply the idea of region ensemble to fingertip localization from stereo images. After feature extraction with two stream DenseNet structured CNN, the feature maps are cropped into nine regions in both left and right streams separately. For each cropping center, the regions of left feature maps and right feature maps are concatenated along the channel dimension to fuse the information of two streams, generating nine concatenated regions. Then the concatenated regions are integrated by FC layers to estimate hand poses.

The labels given in ThuHand17 dataset are pixel coordinates of each joint: (u_l, v_l) and (u_r, v_r), where v_l equals to v_r, which are not consistent with the input of Bi-REN because the input images are crops around the centroid of the hand region but not the entire original images. Therefore, in order to reduce the difficulties of mapping cropped stereo images to some variables corresponding to the positions of joints, we transform the labels into easier mapping forms (Eq.1):

$$label = \begin{pmatrix} \frac{(u_l - c_{xl}) + (u_r - c_{xr})}{w} \\ \frac{2((u_l - c_{xl}) - (u_r - c_{xr}))}{w} \\ \frac{(v_l - c_{yl}) + (v_r - c_{yr})}{h} \end{pmatrix}, \tag{1}$$

where (u_l, v_l) and (u_r, v_r) are the pixel positions of the hand joint in the left image and the right image separately, (c_{xl}, c_{yl}) and (c_{xr}, c_{yr}) are the centroids of the segmented hand region in the left and right images separately, w and h are the width and height of the cropped images (before resized). Note that the label used in our Bi-REN, which is applicable to stereo images, also differs from REN.

Bi-REN minimizes the smooth L1 loss [7] between the output of the last fully connected layer and the transformed label (Eq.1). Afterwards, a pose recovery layer is added to transform the outputs of the last fully connected layer into hand poses (in the form of (u_l, v_l) and (u_r, v_r), with v_l equals to v_r).

3 Preprocessing

In this section, we give a brief introduction of our preprocessing procedure. Since the stereo images in ThuHand17 dataset [21] are captured with the infrared imaging device Leap Motion [1], we can extract hand regions using a thresholding method.

Two different thresholds th_1 and th_2 are set considering the trade-off between preserving the completeness of the entire hand and removing the noise in the background. The larger one (th_1) is first used to obtain rough binary images and only the connected components with maximum area are preserved to calculate the centroids of hand regions. Stereo images are cropped around the centroids. Then we use the same method but with th_2 to get delicate binary images from cropped stereo images. For convenience, the centroids of left and right image are denoted as (c_{xl}, c_{yl}) and (c_{xr}, c_{yr}) respectively (c_{yr} and c_{yl} are set to equal).

Cropped stereo images and cropped mask images are resized into a predefined size $w_p \times h_p$ before being fed into the networks. In cropped stereo images, the pixels outside the hand mask are set to zero, while other pixels remain unchanged.

4 Experiments and Discussions

In this section, we first evaluate our Bi-REN framework on the binocular image dataset ThuHand17 [21]. Then we conduct extensive experiments for ablation study to discuss the effectiveness of different modules in Bi-REN, more specifically, by evaluating the different methods in Table 2.

4.1 Dataset

In ThuHand17, the training set contains about 117K pairs of binocular images of 8 subjects, with 16 basic hand poses and a lot of transitional poses between adjacent basic poses captured by Leap Motion. Seven subjects performed all the 16 basic poses and the transitional poses, while the remainder one mainly performed several kinds of basic poses. The locations of the five fingertips and the wrist are annotated by TrakSTAR [3], and transformed into pixel positions in both left and right images. The test set contains 10K pairs of binocular images (different from training samples) of 2 subjects.

4.2 Experimental Setup

We implement our Bi-REN with Caffe [11]. We use stochastic gradient descent (SGD) with a mini-batch size of 128. The learning rate starts from 0.001 and is divided by 10 every 75000

iterations (about 27 epochs), and the model is trained for total 300000 iterations (about 109 epochs). Besides, we use a weight decay of 0.0005 and a momentum of 0.9.

4.3 Evaluation Metrics

Following [21], the performance is evaluated by two metrics: 1) *average 3D distance error* is computed as the Euclidean distance between the ground truths and the predictions of 3D coordinates in the coordinate system of Leap Motion (in millimeters). The mean of average 3D distance error of all six joints is also presented; 2) *percentage of success frames* is defined as the percentage of frames in which all 3D Euclidean errors of joints are below a variant threshold.

4.4 Comparison with Previous Work

To demonstrate the effectiveness of our Bi-REN, we compare it against previous work, noted as Chen et al. [4] and TSBNet [21]. The average 3D distance error and the percentage of success frames are shown in Figure 2.

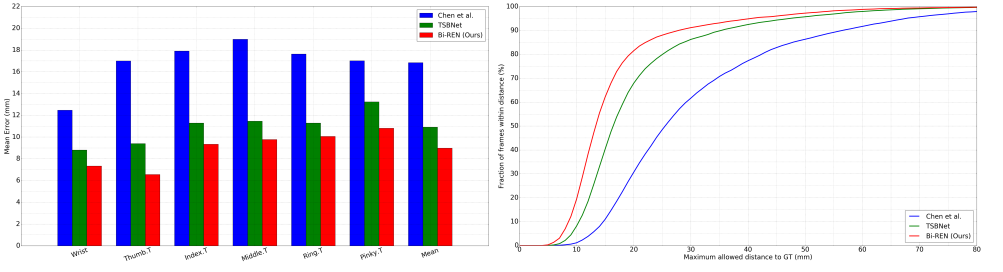

Figure 2: Effect of Bi-REN comparing with Chen et al. [4] and TSBNet [21]. Left: average 3D distance error. Right: percentage of success frames.

As shown in Figure 2, the average errors of each joint estimated with Bi-REN are much smaller than Chen et al. [4] and TSBNet [21]. Except the pinky tip (10.81mm) and the ring tip (10.06mm), the other joints have their mean error smaller than 10mm. The quantitative results of the mean error of all joints (the last histogram labeled as "*Mean*") are presented in Table 1. Compared with Chen et al. [4] and TSBNet [21], the mean error reduced from 16.84mm and 10.91mm to 8.98mm (about 18% improvement in contrast with TSBNet).

The percentage of success frames of Bi-REN outperforms Chen et al. [4] and TSBNet [21] consistently, no matter within a small threshold or a large one. Specifically, under a threshold of 20mm, Bi-REN successfully predicts more than 80% of frames, while TSBNet [21] less than 70% and Chen et al. [4] about 30%. Not only can Bi-REN predict more accurate positions of fingertips and the wrist than previous methods, it also outperforms others in providing high-quality frame predictions under different levels of requirements of accuracy.

Bi-REN runs at 120fps on a NVIDIA Geforce 1080TI GPU during inference, while Chen et al. [4] and TSBNet [21] run at 530fps and 190fps respectively. With some loss of speed but still promising for real-time applications, Bi-REN achieves huge improvement in accuracy.

Method	Mean Error (mm)
Chen et al. [4]	16.84
TSBNet [21]	10.91
Bi-REN (Ours)	**8.98**

Table 1: Quantitative result of mean error of all joints among previous work and our Bi-REN. Bi-REN outperforms others.

4.5 Ablation Study

In order to promote the future research on fingertip localization from stereo images and perform ablation studies, besides Bi-REN, we present several models which are produced by replacing some modules in Bi-REN.

4.5.1 Module Introduction

In this section, we first introduce the substitute modules from Bi-REN, then the different models constructed by concatenating different modules.

(1) Feature Extraction Module. In Bi-REN, we use the structure of DenseNet [9] to extract feature maps from stereo images. For comparison, in another method (Basic-CNN), we use a basic CNN architecture (the baseline in [20] with residual connections, but in two stream style for stereo images) for feature extraction, see Figure 3. We use average pooling in DenseNet structured CNN as [9]. In Basic-CNN, we use max pooling instead, following [20].

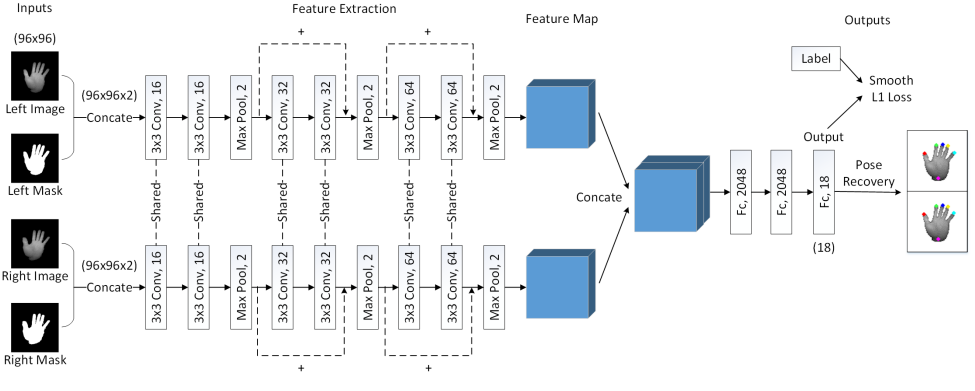

Figure 3: The basic CNN architecture in two streams. "$k \times k$ *Conv*$,c$" represents a convolution layer with a $k \times k$ size kernel and c channels ($stride = 1, pad = 1$). "*Max Pool*$,2$" denotes a max pooling layer with kernel size 2×2 ($stride = 2, no\ padding$). "Fc,n" represents a fully connected layer with n neurons. The numbers inside parentheses denote the sizes of images or vectors.

(2) Region Ensemble Module. REN [20] divides the feature maps into several regions and integrates them in fully connected layers. We apply REN to fingertip localization from stereo images and use it in our Bi-REN. In other methods (Basic-CNN and DenseNet-CNN),

the feature maps of two streams are forwarded into FC layers directly after they are concatenated.

By concatenating different modules, we propose several methods (See Table 2).

Method	Feature Extraction	Regression Architecture
Basic-CNN	Basic CNN	FC layers
DenseNet-CNN	DenseNet	FC layers
Bi-REN	DenseNet	REN

Table 2: The different methods proposed for fingertip localization from stereo images. Basic-CNN is the most basic model (Figure 3) in Section 4.5.1. DenseNet refers to the network containing the first two convolution layers and three dense blocks in Section 2.3. REN refers to region ensemble module in Section 4.5.1. Feature extraction modules of all methods are in two same-structure and shared-parameter streams.

4.5.2 Module Effects

In this section, we explore the effects of each module in Section 4.5.1 by conducting experiments of the methods in Table 2.

(1) Feature Extraction. We test two kinds of feature extraction modules: a shallow CNN in two streams with the same structure and shared parameters and residual connections, and a two-stream DenseNet-CNN. As shown in Figure 4, all the six joints have a smaller mean error using DenseNet-CNN compared with Basic-CNN. The average errors of all joints are 10.20mm using DenseNet-CNN (5.20% better than Basic-CNN, see Table 3). The percentage of success frames is also higher while using DenseNet structure for feature extraction (See Figure 4).

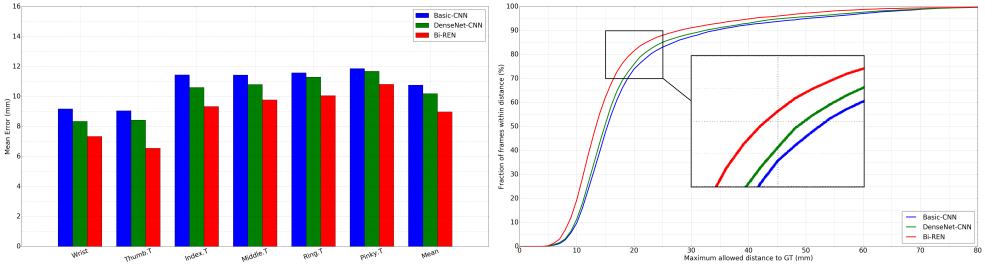

Figure 4: Effect of modules in Bi-REN. Left: average 3D distance error. Right: percentage of success frames.

(2) Region Ensemble. Without region ensemble method, DenseNet-CNN performs poorer than Bi-REN in per joint error, (10.20mm versus 8.98mm, see Table 3 and Figure 4, Bi-REN is 11.96% better). The improvement after using region ensemble architecture (1.22mm, 11.96% improvement from DenseNet-CNN to Bi-REN) is larger than replacing the feature extraction module (0.56mm, 5.20% raises from Basic-CNN to DenseNet-CNN), proving that the region ensemble method has a more significant effect than feature extraction method. As for the percentage of success frames, there is a wide gap between Bi-REN and DenseNet-CNN under different allowed thresholds of distance to the ground truths.

Method	Mean Error (mm)
Basic-CNN	10.76
DenseNet-CNN	10.20
Bi-REN	**8.98**

Table 3: Quantitative result of mean error of all joints among our different models.

Note that our Basic-CNN (6 convolution layers) and DenseNet-CNN (10 convolution layers) also outperforms [4, 7] (12 convolution layers).

4.6 Qualitative Results

Figure 5 shows some examples of the results of Bi-REN. It can be seen that the predictions are very closed to ground truths. The estimation is still promising for some difficult cases, such as several fingers occluded and side viewpoints.

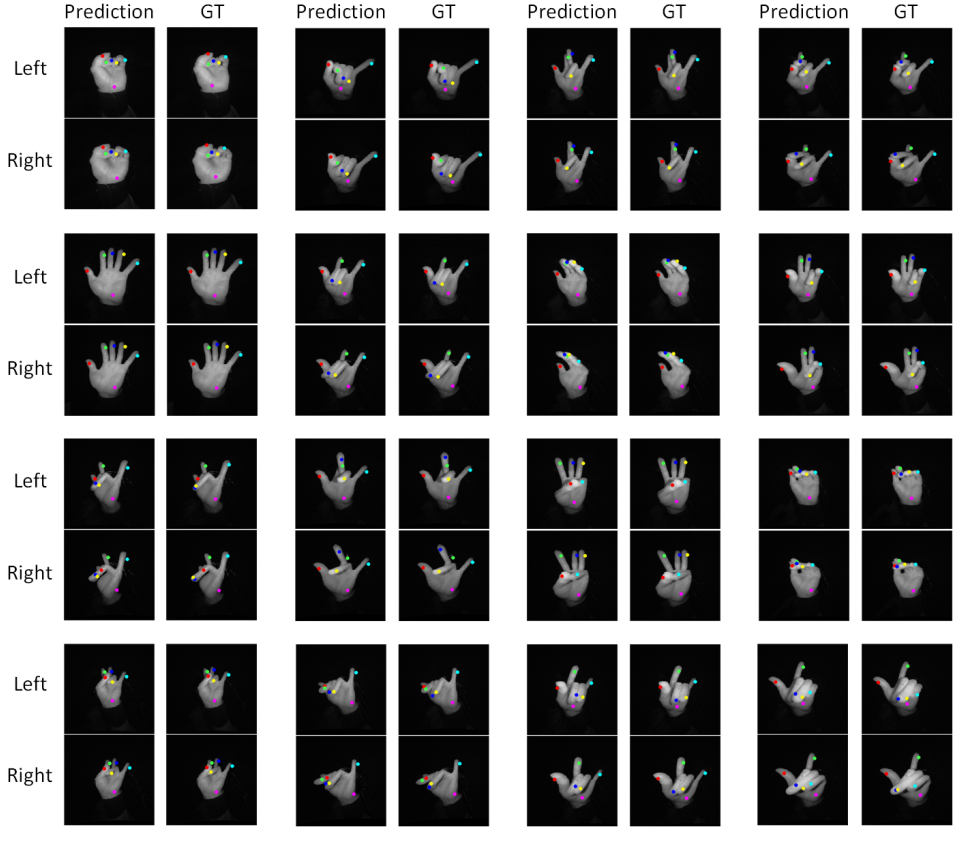

Figure 5: Qualitative results of Bi-REN. Left and right images are shown in adjacent two rows. Bi-REN predictions and the ground truths (GT) are shown in adjacent two columns.

5 Conclusion

In this paper we propose a novel scheme named Bi-REN for fingertip localization from stereo images directly, without conversion into depth images. It first uses DenseNet to extract feature maps, in two streams with the same structure and shared parameters for left and right image separately. Then feature regions are cropped from the feature maps and fused by fully connected layers and concatenation. Experimental results on ThuHand17 binocular hand pose dataset demonstrate the effectiveness of our method. Our future work includes exploring better fusion strategies for the two streams, and building a large dataset since there are few stereo image based datasets publicly accessible and ThuHand17 is not large enough to promote the research on fingertip localization from stereo images.

References

[1] Leap Motion. https://www.leapmotion.com/.

[2] ThuHand17. https://sites.google.com/view/thuhand17.

[3] Ascension Trakstar. http://www.ascension-tech.com/.

[4] Xinghao Chen, Guijin Wang, and Hengkai Guo. Accurate fingertip detection from binocular mask images. In *Visual Communications and Image Processing (VCIP), 2016*, pages 1–4. IEEE, 2016.

[5] Xinghao Chen, Guijin Wang, Hengkai Guo, and Cairong Zhang. Pose guided structured region ensemble network for cascaded hand pose estimation. *Neurocomputing*, 2018.

[6] Liuhao Ge, Hui Liang, Junsong Yuan, and Daniel Thalmann. 3D convolutional neural networks for efficient and robust hand pose estimation from single depth images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, page 5, 2017.

[7] Ross Girshick. Fast R-CNN. In *Computer Vision (ICCV), 2015 IEEE International Conference on*, pages 1440–1448. IEEE, 2015.

[8] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.

[9] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, volume 1, page 3, 2017.

[10] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015.

[11] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014.

[12] Gyeongsik Moon, Ju Yong Chang, and Kyoung Mu Lee. V2V-PoseNet: Voxel-to-voxel prediction network for accurate 3D hand and human pose estimation from a single depth map. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2018.

[13] Chenbo Shi, Guijin Wang, Xuanwu Yin, Xiaokang Pei, Bei He, and Xinggang Lin. High-accuracy stereo matching based on adaptive ground control points. *IEEE Transactions on Image Processing*, 24(4):1412–1423, 2015.

[14] Xiao Sun, Yichen Wei, Shuang Liang, Xiaoou Tang, and Jian Sun. Cascaded hand pose regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 824–832, 2015.

[15] James S Supancic, Grégory Rogez, Yi Yang, Jamie Shotton, and Deva Ramanan. Depth-based hand pose estimation: data, methods, and challenges. In *Proceedings of the IEEE international conference on computer vision*, pages 1868–1876, 2015.

[16] Danhang Tang, Hyung Jin Chang, Alykhan Tejani, and Tae-Kyun Kim. Latent regression forest: Structured estimation of 3D articulated hand posture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3786–3793, 2014.

[17] Jonathan Tompson, Murphy Stein, Yann Lecun, and Ken Perlin. Real-time continuous pose recovery of human hands using convolutional networks. *ACM Transactions on Graphics (ToG)*, 33(5):169, 2014.

[18] Chengde Wan, Thomas Probst, Luc Van Gool, and Angela Yao. Crossing nets: Combining GANs and VAEs with a shared latent space for hand pose estimation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[19] Guijin Wang, Xuanwu Yin, Xiaokang Pei, and Chenbo Shi. Depth estimation for speckle projection system using progressive reliable points growing matching. *Applied optics*, 52(3):516–524, 2013.

[20] Guijin Wang, Xinghao Chen, Hengkai Guo, and Cairong Zhang. Region ensemble network: Towards good practices for deep 3D hand pose estimation. *Journal of Visual Communication and Image Representation*, 2018.

[21] Yi Wei, Guijin Wang, Cairong Zhang, Hengkai Guo, Xinghao Chen, and Huazhong Yang. Two-stream binocular network: Accurate near field finger detection based on binocular images. In *Visual Communications and Image Processing (VCIP), 2017 IEEE*, pages 1–4. IEEE, 2017.

[22] Shanxin Yuan, Guillermo Garcia-Hernando, Björn Stenger, Gyeongsik Moon, Ju Yong Chang, Kyoung Mu Lee, Pavlo Molchanov, Jan Kautz, Sina Honari, Liuhao Ge, et al. Depth-based 3D hand pose estimation: From current achievements to future goals. In *IEEE CVPR*, 2018.

[23] Cairong Zhang, Guijin Wang, Hengkai Guo, Xinghao Chen, Fei Qiao, and Huazhong Yang. Interactive hand pose estimation: Boosting accuracy in localizing extended finger joints. *Electronic Imaging*, 2018(2):251–1–251–6, 2018. ISSN 2470-1173. doi: doi:10.2352/ISSN.2470-1173.2018.2.VIPC-251.

[24] Jiawei Zhang, Jianbo Jiao, Mingliang Chen, Liangqiong Qu, Xiaobin Xu, and Qingxiong Yang. 3D hand pose tracking and estimation using stereo matching. *arXiv preprint arXiv:1610.07214*, 2016.