NTIRE 2021 Depth Guided Relighting Challenge Factsheet

Zhipeng Luo, Zhiguang Zhang, Janye He

March 20, 2021

1 Team details

- Team name

  DeepBlueAI

- Team leader name

  Zhipeng Luo

- Team leader address, phone number, and email

  27 Zhongguancun Street, Haidian District, Beijing, China;

  +86 13051510090;

  luozp@deepblueai.com

- Rest of the team members

  Zhiguang Zhang; Jianye He

- Team website URL (if any)

  None

- Affiliation

  DeepBlue Technology Shanghai Co.,Ltd

- Affiliation of the team and/or team members with NTIRE2021 sponsors (check the workshop website)

  No affiliation with the sponsors

- User names and entries on the NTIRE2021 Codalab competitions (development/validation and testing phases)

  DeepBlueAI

  Track1: 0/7/8

  Track2: 0/7/10

- Best scoring entries of the team during development/validation phase

  Track1: PSNR: 18.4800 SSIM: 0.6757;

  Track2: PSNR: 20.0637 SSIM: 0.7196

- Link to the codes/executables of the solution(s)

  codes

- Link to the restoration results of all frames

  Track1 Test Result

  Track2 Test Result

2 Contribution details

- Title of the contribution

  An innovative way for depth guided image relighting

- General method description

  In our method, based on U-Net, we designed a new depth guide image relighting network with RGB-D image as input and output the relighted image.

  In the encoder network, we designed a fusion network to extract features of the input and guide RGB-D images. Then use a lighting estimation net to estimate the Temperature and lighting direction of the guide image, and a lighting to feature net to encoder the result and replace the feature of the input image. To better recognition the temperature and lighting direction of input and guide image, we designed and add Dilated Residual Network(DRN) module to the encoder network at each level. The model was trained with loss MSE, SSIM Loss, GradLoss, Focal Loss, LPIPS Loss.

  In the test stage, for track2 we just input the input and guide RGB-D images to the network and get output. For track1, we use the network and weight data trained in track2 and select one image(East, 4500K) in the training dataset as guide to get the result for track1.

- References

  1. Zhou H , Hadap S , Sunkavalli K , et al. Deep Single-Image Portrait Relighting[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019.

  2. Puthussery D , Sethumadhavan H P , Kuriakose M , et al. WDRN: A Wavelet Decomposed RelightNet for Image Relighting[M]// Computer Vision – ECCV 2020 Workshops, Glasgow, UK, August 23–28, 2020, Proceedings, Part III. 2021.

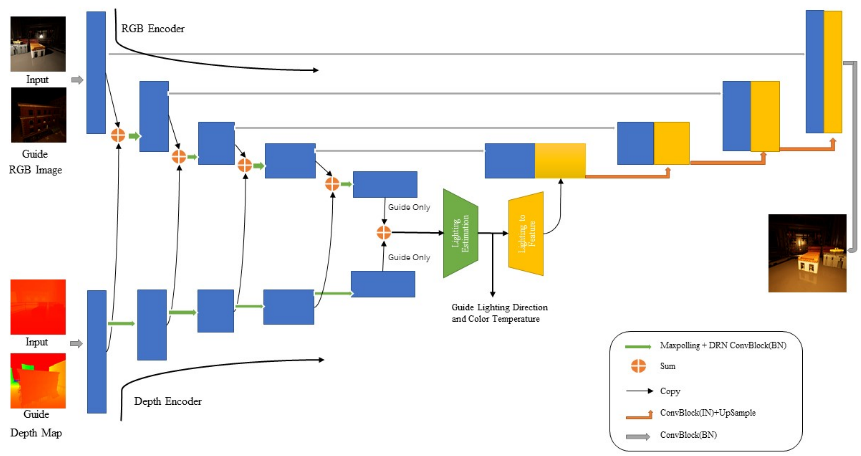

Figure 1: The pipeline of our method.

3. Chen X , Lin K Y , Wang J , et al. Bi-directional Cross-Modality Feature Propagation with Separation-and-Aggregation Gate for RGB-D Semantic Segmentation[J]. 2020.

4. Hu, Z., Huang, X., Li, Y., Wang, Q.: SA-AE for any-to-any relighting. In: Pro-ceedings of the European Conference on Computer Vision Workshops (ECCVW)(2020).

5. Hazirbas C , Ma L , Domokos C , et al. FuseNet: Incorporating Depth into Semantic Segmentation via Fusion-Based CNN Architecture[J]. 2016.

6. Chen L C , Papandreou G , Kokkinos I , et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848.

7.Chen X , Lin K Y , Wang J , et al. Bi-directional Cross-Modality Feature Propagation with Separation-and-Aggregation Gate for RGB-D Semantic Segmentation[J]. 2020.

- Representative image / diagram of the method(s)

  Shown as Figure1 and Figure2.

3 Global Method Description

- Total method complexity: all stages

  Total parameters: 24.62MB

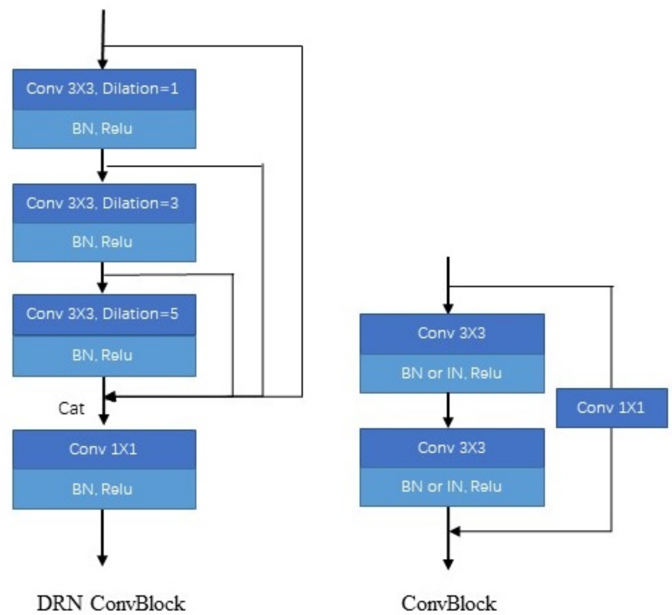

Figure 2: DRN ConvBlock and ConvBlock.

Average flops cost: 195226.05MB

- Which pre-trained or external methods / models have been used (for any stage, if any)

  None

- Which additional data has been used in addition to the provided NTIRE training and validation data (at any stage, if any)

  None

- Training description

  We trained the model with 2*V100 32 GiB memory, with an input size of 512x512 and mini-batches of 16 for 200 epochs. Adam optimizer with initial learning rate 1e-4, and decayed by a factor of 0.1 after every 80 epochs. Especially, we did not use any model initialize method.

  To get better performance, firstly we train the model on the training set with Cross Entropy Loss, MSE, SSIM Loss, and select the best performance epoch model weight as pre-trained model. Then Re-optimize the model with Focal Loss, MSE, SSIM Loss, LPIPS Loss or Gradients Loss. Especially, we did not use any data augmentation.

Table 1: Track2 results on validation dataset.

Method	PSNR	SSIM
SA-AE[4]	18.0695	0.6480
Ours	20.0637	0.7196

- Testing description

  In the test stage, use input with the size of 512x512 and get output with the size of 512x512, then upscale the output to 1024x1024.

- Quantitative and qualitative advantages of the proposed solution

  Our method is a user-controlled relighting method, it can work with arbitrary guide images, even if a guide image is missing, we can still perform relighting according to user demands.

- Results of the comparison to other approaches (if any)

- Results on other benchmarks (if any)

  None

- Novelty degree of the solution and if it has been previously published

  Based on some previous works related to semantic segmentation, object detection, image relighting we proposed our method and have never been published.

  In our methods, we designed a new depth guided relighting network, and designed Dilated Residual Network, which brings the greatest performance improvement.

- It is OK if the proposed solution is based on other works (papers, reports, Internet sources (links), etc). It is ethically wrong and a misconduct if you are not properly giving credits and hide this information.

4 Competition particularities

Any particularities of the deployed solution for the competition (if applicable)

First, based on U-Net we designed a depth guided relighting network which use RGB-D image as input and output the relighted image. The encoder network has two paths, one for RGB image, one for Depth map, and we proposed a method to fusion RGB image feature and Depth map feature.

Second, based on some previous methods, like SA-AE[4], we designed and add a lighting-color-estimate net and a lighting-to-feature net, the difference is we completely replace the feature of input image with the lighting-to-feature output of the guide image. And use Focal loss as the loss function.

Table 2: Track2 results on validation dataset with fused methods.

Method	**PSNR**	**SSIM**
baseline	19.2043	0.7003
fuised method	20.0637	0.7196

Third, we designed and add Dilated Residual Network(DRN) module to the encoder network at each level.

Forth, in different convolution blocks, we use different normalization methods. In the DRN convolution block, we use Batch Normalization, in the convolution blocks of the decoder network we use Instance Normalization, others are Batch Normalization.

Fifth, without use data augmentation, we can also achieve a good score.

5 Ensembles and fusion strategies

- Describe in detail the use of ensembles and/or fusion strategies (if any).

  Average the outputs predict by different model.

- What was the benefit over the single method?

  Get better PSNR and SSIM, and look better.

- What were the baseline and the fused methods?

  Shown as Table 2.

6 Technical details

- Language and implementation details (including platform, memory, parallelization requirements)

  Python3.8

  Pytorch1.6

  Training: 2*V100 32G

  Testing: 1*V100 16G

- Human effort required for implementation, training and validation?

  No

- Training/testing time? Run time at test inference per image.

  Training time about 36h, 0.033s at test per image

- Comment the robustness and generality of the proposed solution(s). Is it easy to deploy it for other data domains, other illuminants and illumination angles?

  Our method is a user-controlled relighting method, it can work with arbitrary guide images, even if a guide image is missing, we can still perform relighting according to user demands.

- Comment the efficiency of the proposed solution(s).

  Our method is an End-to-End method, and we can get what we want easily.

7 Other details

- Planned submission of a solution(s) description paper at NTIRE 2021 workshop.

  It may be decided based on the ranking of the competition. If we have achieved good rank, we will submit a solution description paper.

- General comments and impressions of the NTIRE2021 challenge.

- What do you expect from a new challenge in image restoration, enhancement and manipulation?

- Other comments: encountered difficulties, fairness of the challenge, proposed subcategories, proposed evaluation method(s), etc.