Places Challenge 2017

Scene Parsing Task

CASIA_IVA_JD

Jun Fu, Jing Liu, Longteng Guo, Haijie Tian, Fei Liu, Hanqing Lu

Yong Li, Yongjun Bao, Weipeng Yan

National Laboratory of Pattern Recognition,
Institute of Automation,
Chinese Academy of Sciences

Model Team,
Business Growth BU,
JingDong (JD)

Contents

- Data analysis
- Stacked Deconvolutional Network (SDN)
- Ensemble modeling

Data analysis

Basic information of the data

- 20210 images for training, and 2000 images for validation, 3352 images for testing
- 150 labels including 35 stuff concepts and 115 discrete objects

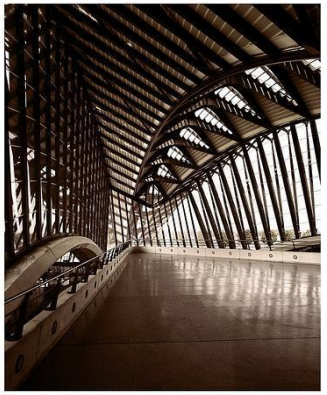
Further statistics

- For each label of training data:
  The number of labeled images: 42(escalator) ~ 11664(wall)
- For each image of training data :
  The number of labels: 0 ~ 31 average: 8.17
- Image width and height:
  Training data: min size: 96 x 130 max size: 2100 x 2100
  Validation data: min size: 200 x 200 max size: 1600 x 1600

Data analysis

Challenge

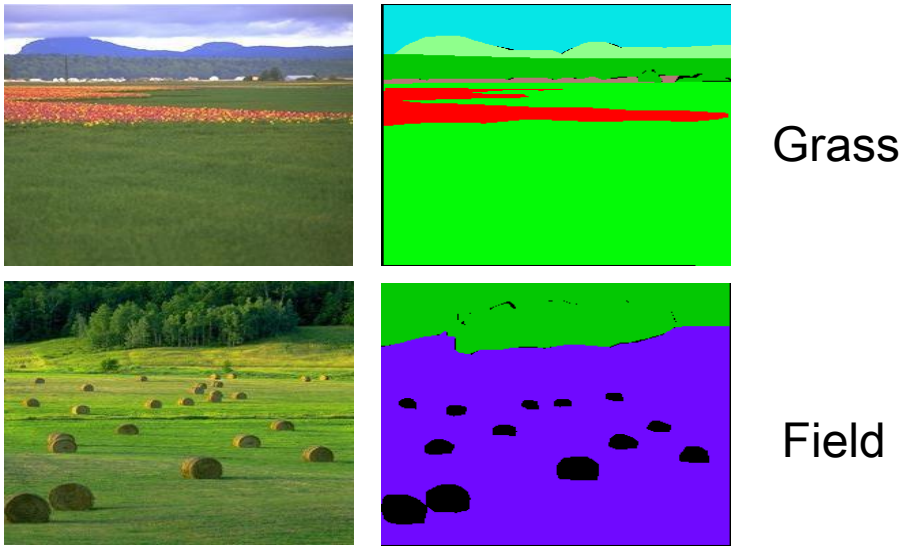
- Diverse and complex scenes
  - Contains various objects in some scenes
  - Background clutter, light condition change, deformation,…



Data analysis

Challenge

- Diverse and complex images
- Similar semantic label
  - Field/Earth/Grass, Desk/Table, Mountain/Hill, ……



Grass

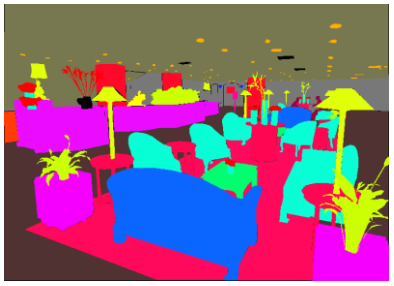
Field

Data analysis

Challenge

- Diverse and complex images
- Similar semantic label
- Multi-scale information
  - Image size
  - Stuff and objects in images



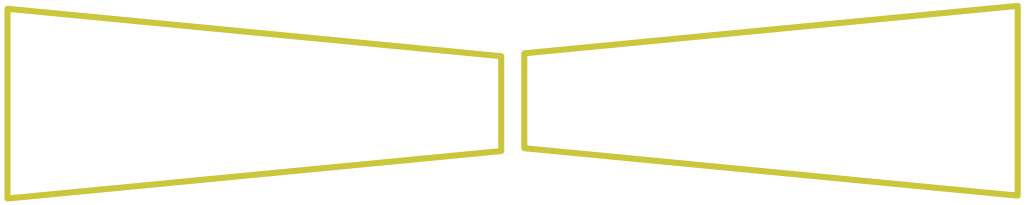



Contents

- Data analysis
- Stacked Deconvolutional Network (SDN)
- Ensemble modeling

Related Work

Deconvolutional network



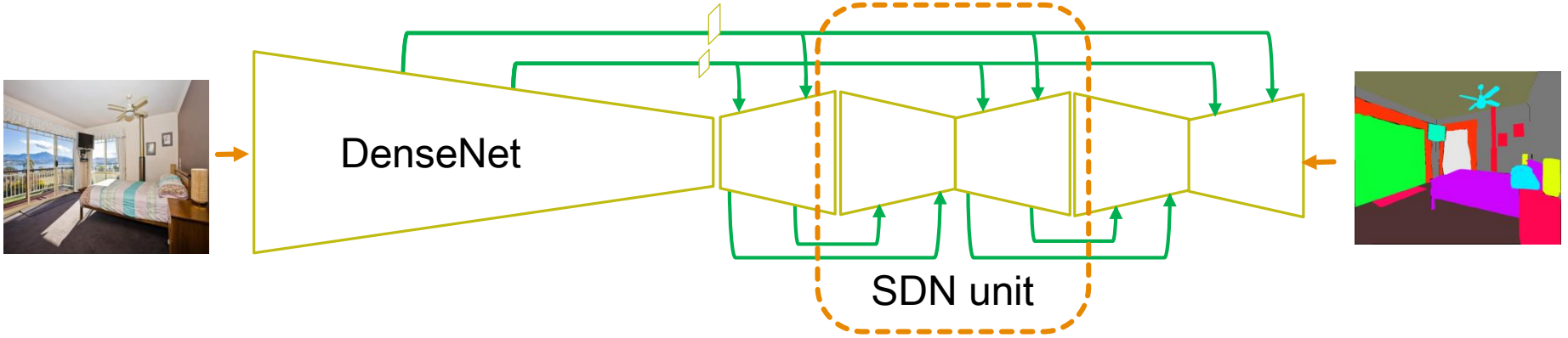
- Common network structure for pixel-level vision task
- Encoder module: capture context
- Decoder module: recover spatial information
- DeconvNet, SegNet, Light-DCNN

Drawbacks

- Limited learning ability (VGG)
- Difficulty in training

Stacked Deconvolutional Network

The architecture of stacked deconvolutional network

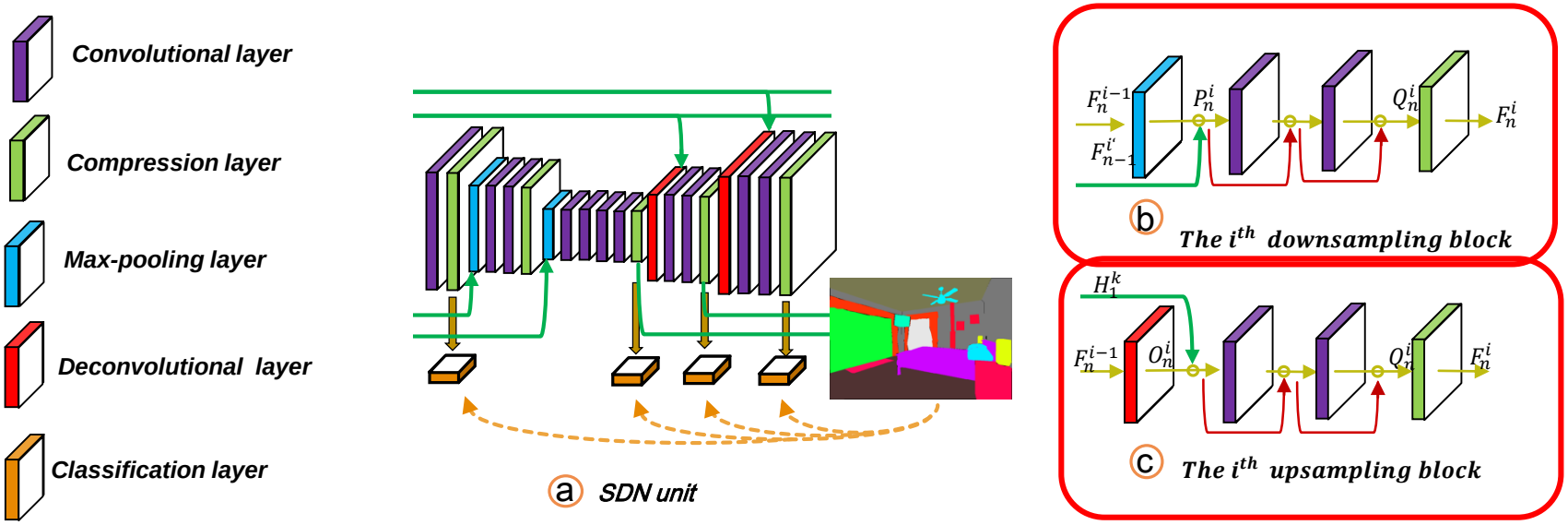


- Design an efficient shallow deconvolutional network (called as SDN unit), stack multiple SDN units one by one with dense connections
- Other designs:
  - Intra-unit connections
  - inter-unit connections
  - hierarchical supervision

SDN link: https://arxiv.org/pdf/1708.04943.pdf

Stacked Deconvolutional Network

The architecture of SDN unit

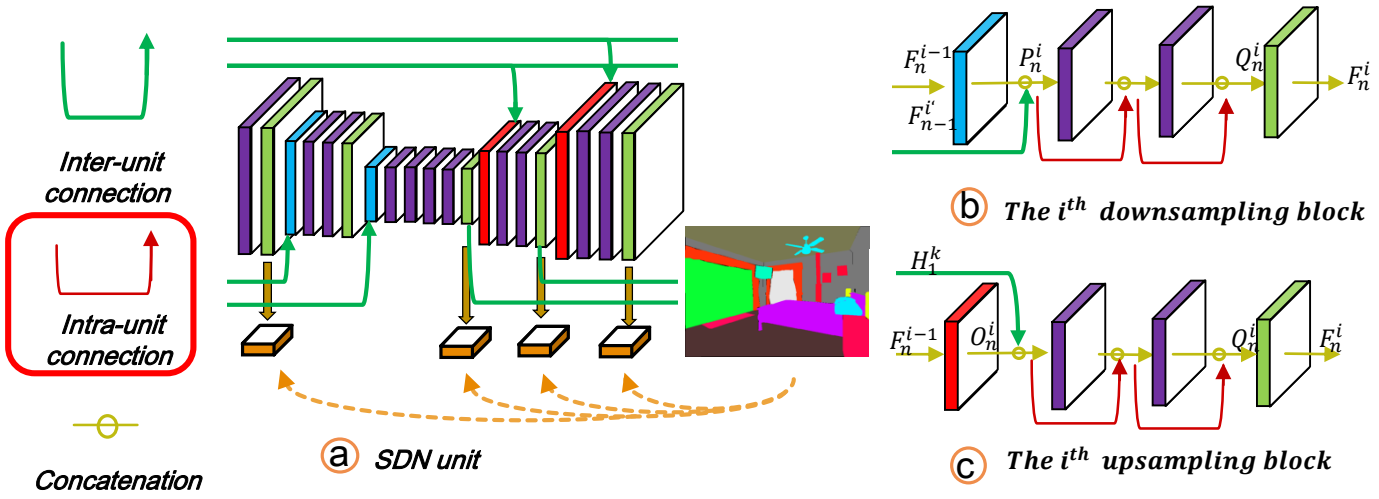


- Encoder module: two downsampling blocks
  - Enlarge the receptive fields of the Network
- Decoder module: two upsampling blocks
  - Achieve a more refined reconstruction of the feature maps

SDN link: https://arxiv.org/pdf/1708.04943.pdf

Stacked Deconvolutional Network

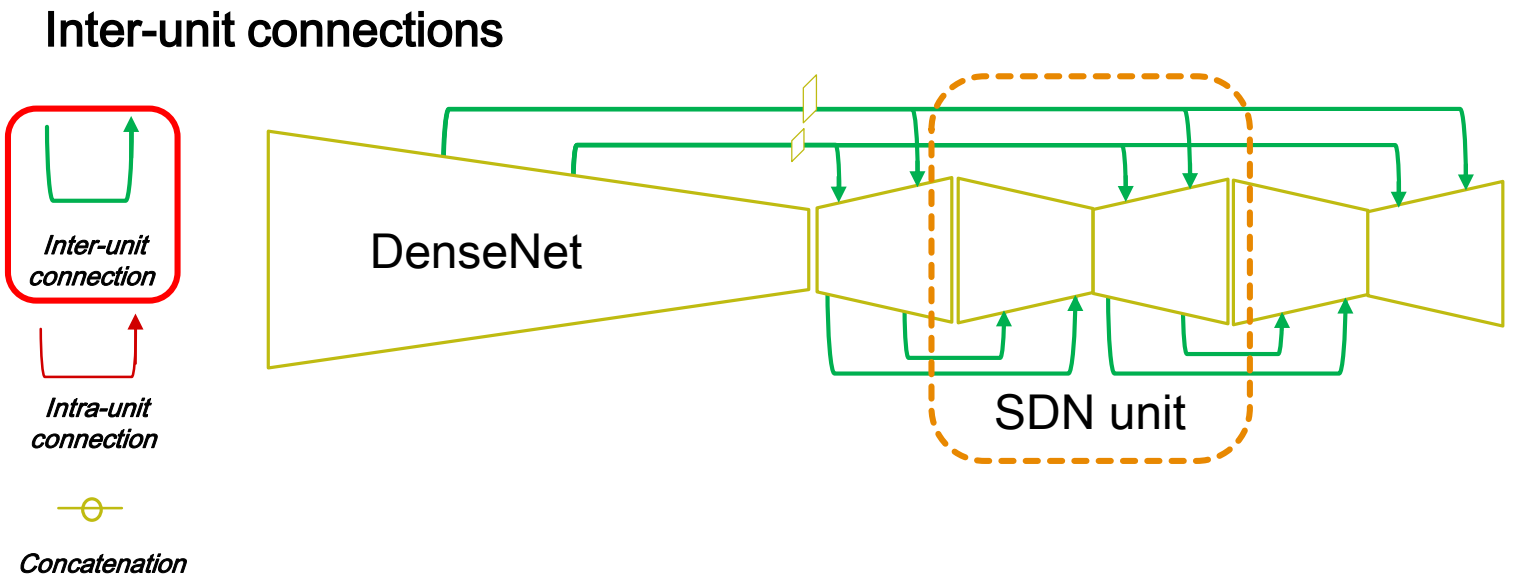
Intra-unit connections



- Intra-unit connections
  - Dense connections in a downsampling/upsampling block
  - Beneficial to the flow of information and gradient propagation throughout the network

SDN link: https://arxiv.org/pdf/1708.04943.pdf

Stacked Deconvolutional Network

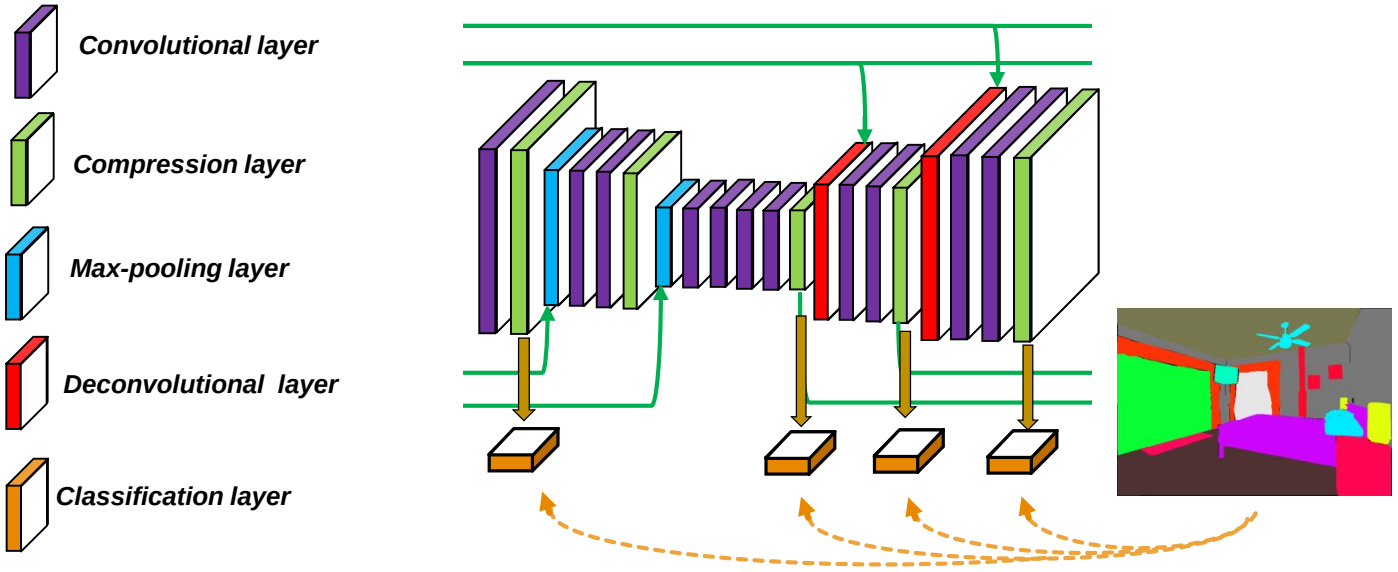


- Inter-unit connections
  - Reuse the multi-scale information across different units
  - Two types of inter-unit skip connections
    - Between any two adjacent SDN units
    - The skip connections from the first SDN units to others

SDN link: https://arxiv.org/pdf/1708.04943.pdf

Stacked Deconvolutional Network

Hierarchical supervision



- Hierarchical supervision
  - Assist training
  - Guarantee the discrimination of the feature maps

SDN link: https://arxiv.org/pdf/1708.04943.pdf

Stacked Deconvolutional Network

Some Training settings:

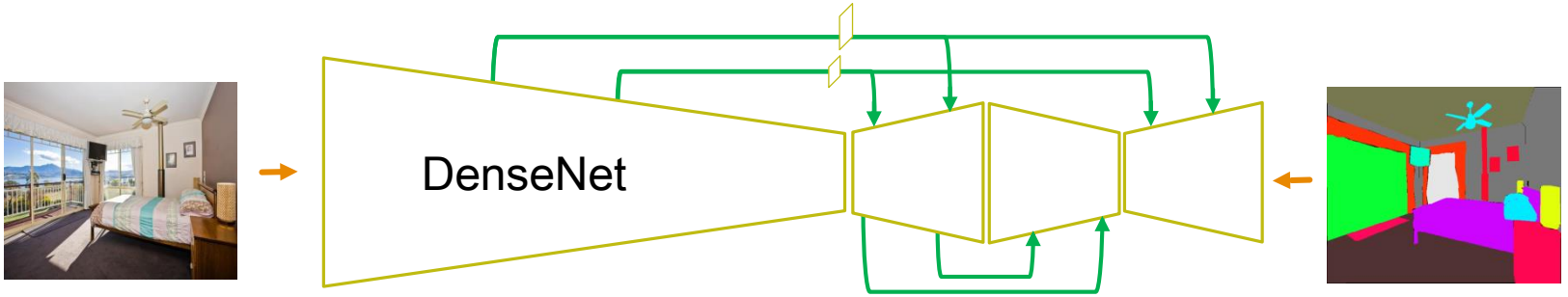
- Data augmentation
  - ✓ scale ratio augmentation (s=[0.5 0.75 1 1.25 1.5])
    $W' = W \cdot s$; $H' = H \cdot s$
  - ✓ aspect ratio augmentation (a=[0.85 1 1.15])
    $W' = W/a$; $H' = H \cdot a$
  - ✓ resize and random crop
- Large cropsize
- Proper learning rate 2.5e-4 and iteration number 100K

Testing scheme:

- Resize the image and testing with sliding window crop
- Multi scale test

Stacked Deconvolutional Network

SDN_M2



SDN_M2 result by mean IoU / pixel accuracy

- Val Data: 44.57/81.22

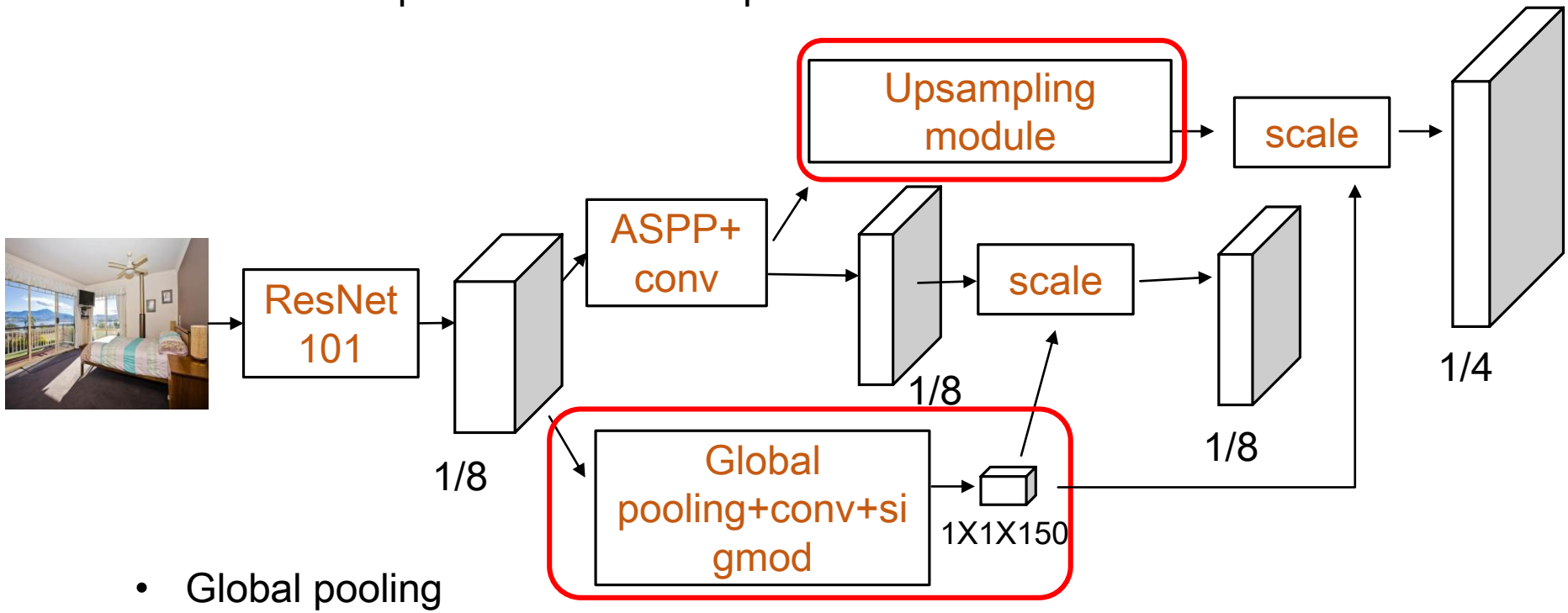
Contents

- Data analysis
- Stacked Deconvolutional Network (SDN)
- Ensemble modeling

Ensemble modeling

Deeplabv3+

- Some improvements on deeplabv3



- Global pooling
- Upsampling module: similar to RefineNet

Deeplabv3+ result Val Data: 44.25/81.02

Ensemble modeling

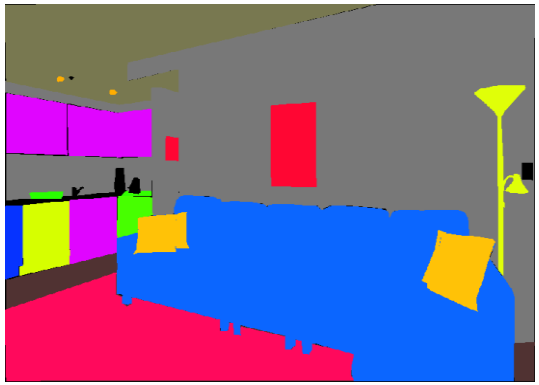
- SDN: **44.57/81.22**
- Deeplabv3+: 44.25/81.02
- ResNet38: 44.07/81.07

By averaging the results of these models,
the score increased to **46.59/82.23**

Deeplabv3+ and ResNet38 adopt training settings and test scheme as the same as SDN, and initialize networks with the models pretrained on ImageNet.

Ensemble modeling

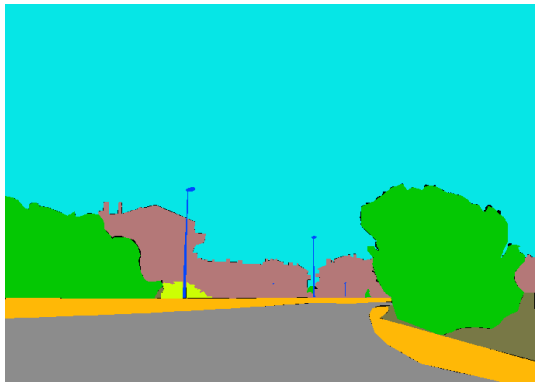
Some scene parsing results



wall	floor
ceiling	cabinet
table	painting
sofa	rug
lamp	cushion
chest of drawers	sink
stove	light
oven	dishwasher

Ensemble modeling

Some scene parsing results

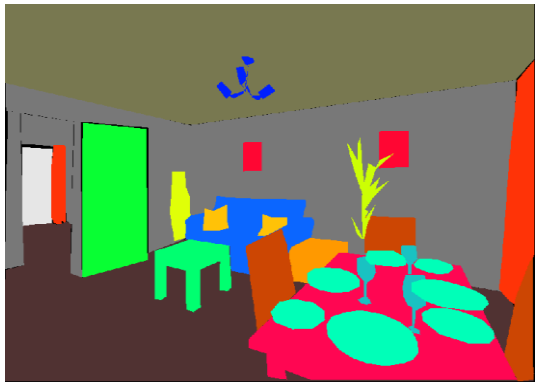


Ensemble modeling

Some scene parsing results

Ensemble modeling

Some scene parsing results



wall	floor
ceiling	windowpane
door	table
plant	curtain
chair	painting
sofa	rug
lamp	cushion
coffee table	light

Contents

- Data analysis
- Stacked Deconvolutional Network (SDN)
- Ensemble modeling

Thanks

Any questions, please contact the author of the work

Email: jliu@nlpr.ia.ac.cn