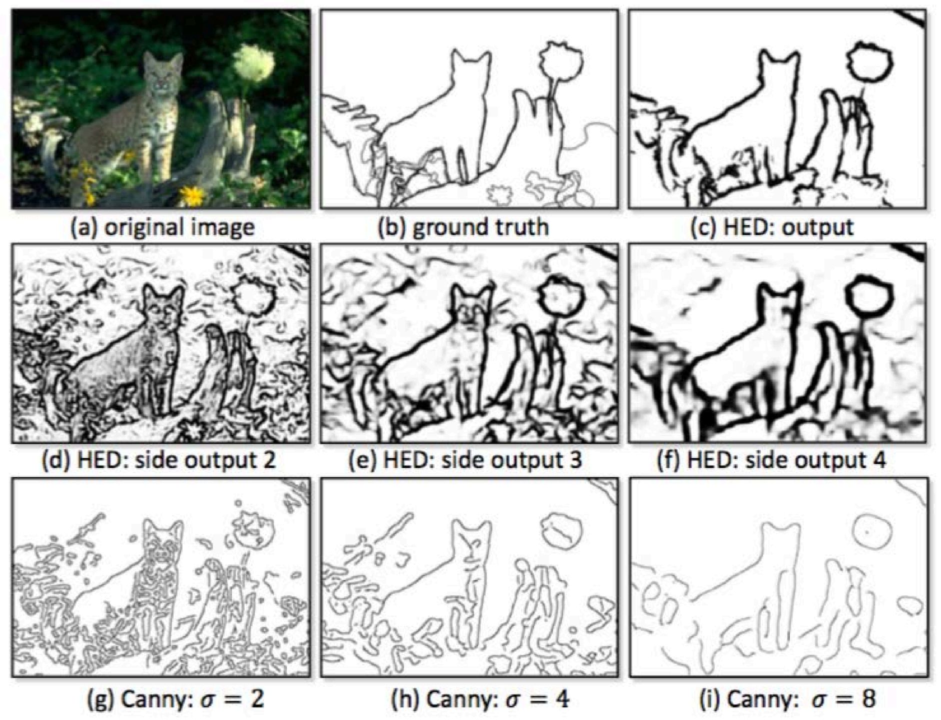
Holistically-Nested Edge Detection (HED)

Mingze Xu & Hanfei Mei
School of Informatics and Computing, Indiana University

Motivation

- **Edge detection** is both fundamental and important to other computer vision areas, such as segmentation, classification, recognition and 3D reconstruction.
- This paper constructs a new **deep learning** model in a fully convolutional neural network which processes the edge detection in an "image-to-image" method.
- This paper uses **multi-scale** and **multi-level** structure to generate 4 side outputs which together contribute to the final fusion result advanced to state-of-art accuracy.



- Figure 1.
- In the first row: (a) shows an example test image in the BSD500 dataset; (b) shows its corresponding edges as annotation by human subjects; (c) displays the HED results.
- In the second row: (d), (e), and (f) respectively show side edge responses from layers 2, 3, and 4 of convolutional neural networks.
- In the third row: (g), (h), and (i) respectively show edge responses from the Canny detector at scales $\sigma = 2.0, \sigma = 4.0$, and $\sigma = 8.0$.
- HED result shows a clear advantage in consistency over Canny.
- Note: Figure reprinted from [1] Saining and Zhouwen.

Multi-scale & Multi-level NN

- **Holistically-nested network** is a relatively simple variant that is able to produce predictions from multiple scales.
- The architecture comprises a single-stream deep network with multiple side outputs. This architecture resembles several previous works, particularly the deeply-supervised net approach which shows that hidden layer supervision can improve both optimization and generalization for image classification tasks.
- The **multiple side outputs** also give us the flexibility to add an additional fusion layer if a unified output is desired.

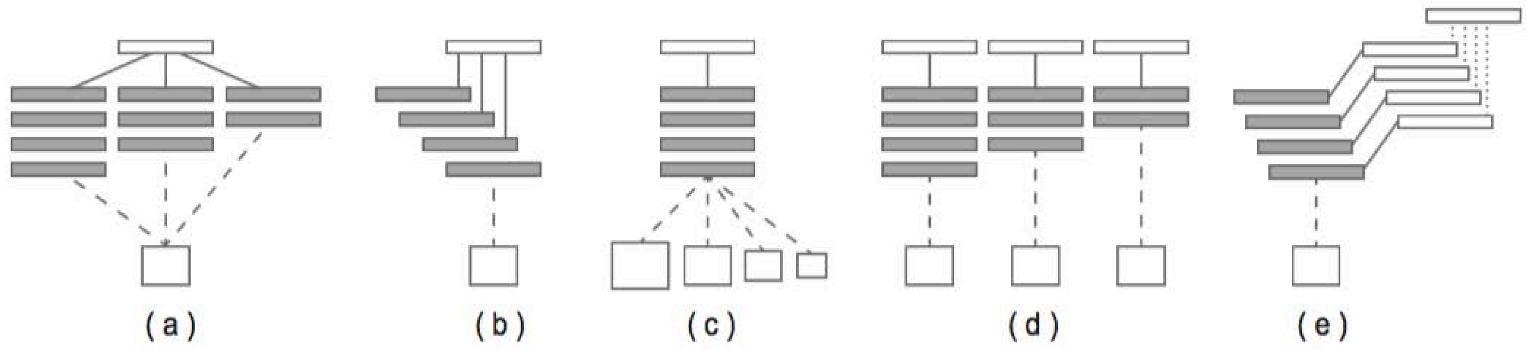


Figure 2. (a) multi-stream architecture; (b) skip-layer net architecture; (c) a single model running on multi-scale inputs; (d) separate training of different networks; (e) **holistically-nested architecture**.
Note: Figure reprinted from [1] Saining and Zhouwen.

Network Architecture

- This paper adopts the VGGNet architecture but makes the following modifications:
- Connecting the side output layer to the last convolutional layer in each stage. The receptive field size of each of these convolutional layers is identical to the corresponding side-output layer;
- Cutting the last stage of VGGNet, including the 5th pooling layer and all the fully connected layers;
- Final HED network architecture has 5 stages, with different strides and different receptive field sizes, which enable the fusion result to recognize reasonable edges in every scale.

layer	**c1_2**	p1	**c2_2**	p2	**c3_3**
rf size	**5**	6	**14**	16	**40**
stride	**1**	2	**2**	4	**4**
layer	p3	**c4_3**	p4	**c5_3**	p5
rf size	44	**92**	100	**196**	212
stride	8	**8**	16	**16**	32

Table 1. The receptive field and stride size in VGGNet used in HED. The **bolded convolutional layers** are linked to additional side-output layers. Note: Table reused from [1] Saining and Zhouwen.

- Given image X, we obtain edge predictions from both the side output layers and the weighted-fusion layer:

$$(\hat{Y}_{\text{fuse}}, \hat{Y}_{\text{side}}^{(1)}, \ldots, \hat{Y}_{\text{side}}^{(M)}) = \text{CNN}(X, (\mathbf{W}, \mathbf{w}, \mathbf{h})^{\star}) \quad (1)$$

$$\hat{Y}_{\text{HED}} = \text{Average}(\hat{Y}_{\text{fuse}}, \hat{Y}_{\text{side}}^{(1)}, \ldots, \hat{Y}_{\text{side}}^{(M)}) \quad (2)$$

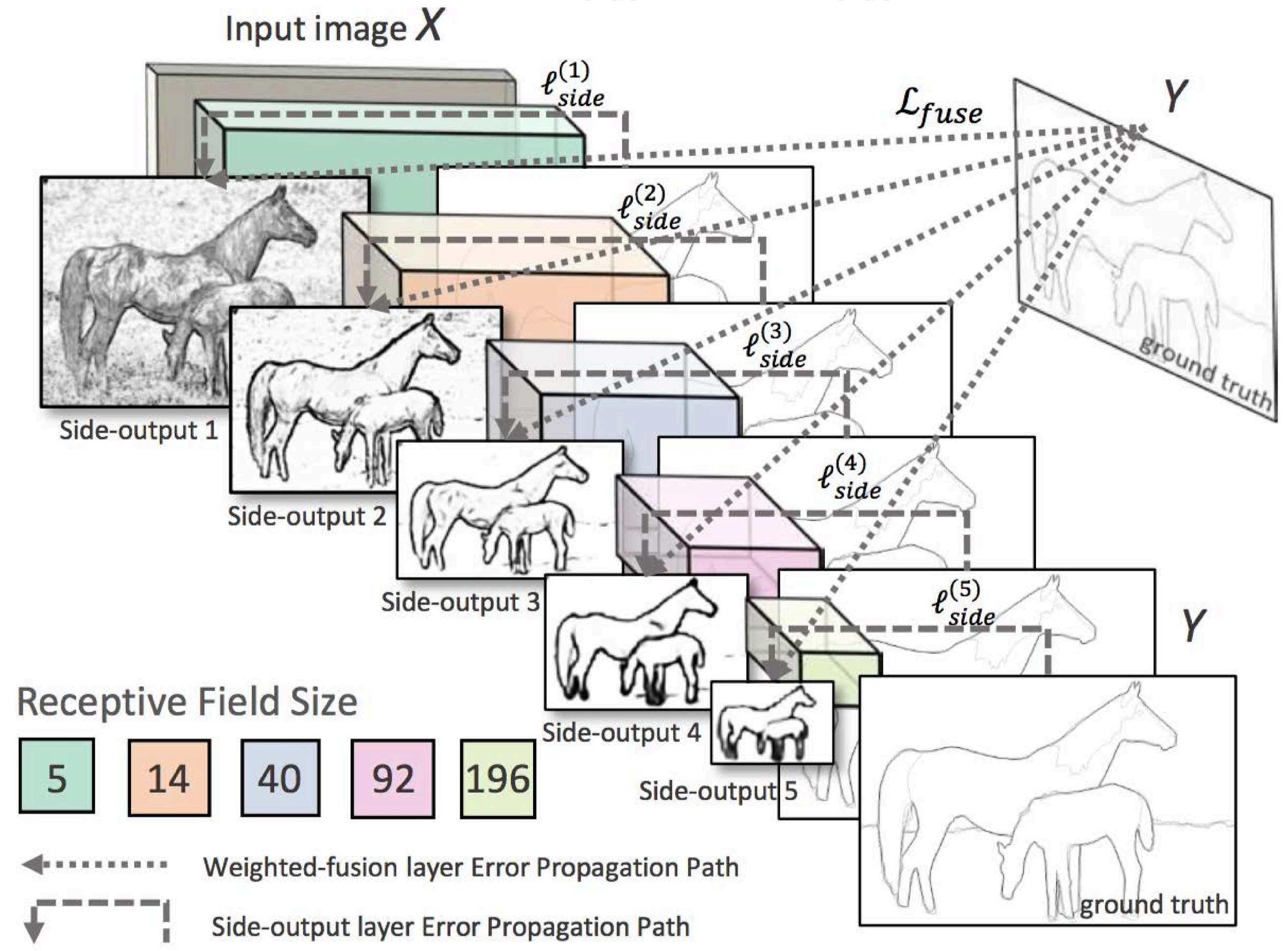


Figure 3. Illustration of the network architecture. The outputs of HED is multi-scale and multi-level, with the side-output-plane size becoming smaller and the receptive field size becoming larger. It also highlights the error back propagation path. Note: Figure reused from [1] Saining and Zhouwen.

Results

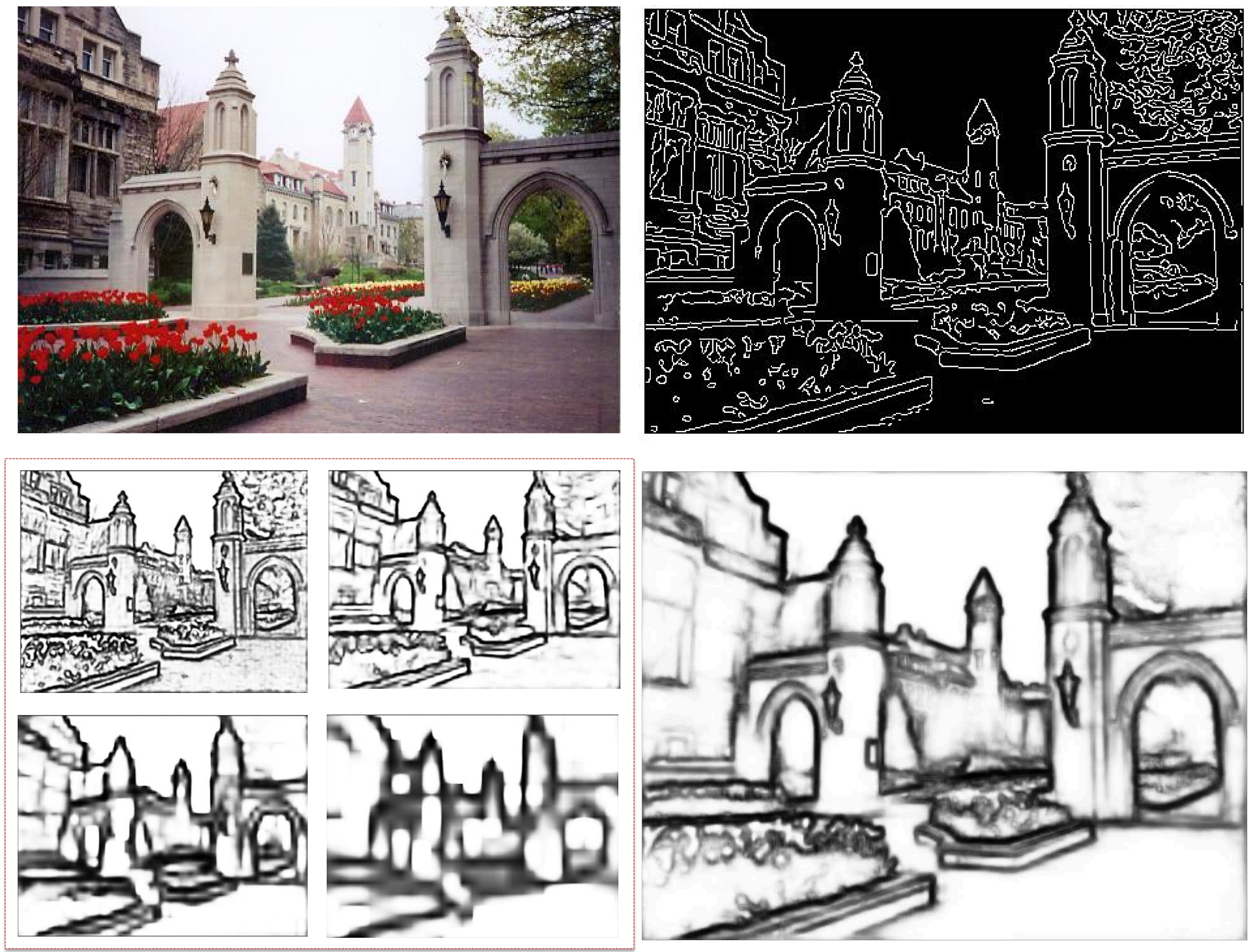


Figure 4.
First line: left) original image; right) Canny output. Second line: left) example side outputs; right) HED output.

Applying to Existing Problem

- In B657's homework 1, we tried to solve Optical Music Recognition (OMR) problem by matching edge maps between templates and an input image.
- In this project, we will use this state-of-art edge detection method to see if better edge maps could help improving the performance of a recognition problem.

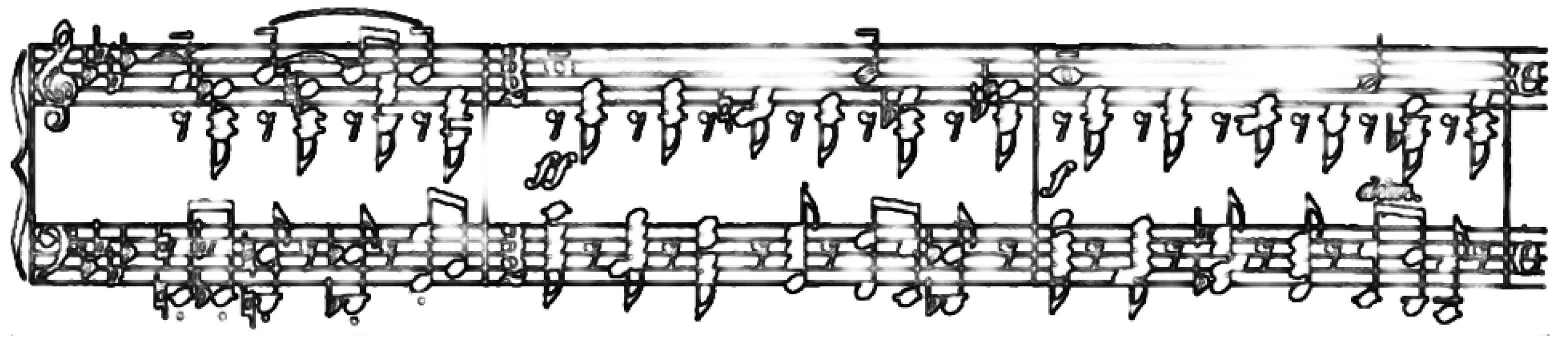


Figure 5. HED edge detection of music optical character image.

- Unfortunately, the result of the experiment shows that using HED edge detection doesn't help too much in OMR problem. The result is almost the same as using Canny edges.

[1] Saining Xie and Zhouwen Tu, "Holistically-Nested Edge Detection," in Proc. ICCV, 2015.
[2] K. Silberman, A. Zisserman. "Very deep convolutional networks for large-scale image recognition," In ICLR, 2015
[3] Liang Chen, Kun Duan, "MIDI-Assisted Egocentric Optical Music Recognition", WACV, 2016.

Acknowledgments: *Texts of HED architecture descriptions are reused from [1] Saining and Zhouwen. And, we gratefully thank Prof. David Crandall for helping improve the quality of this poster.*