Deep Learning for Fluoroscopic Feature Detection

Background Presentation

Liujiang Yan

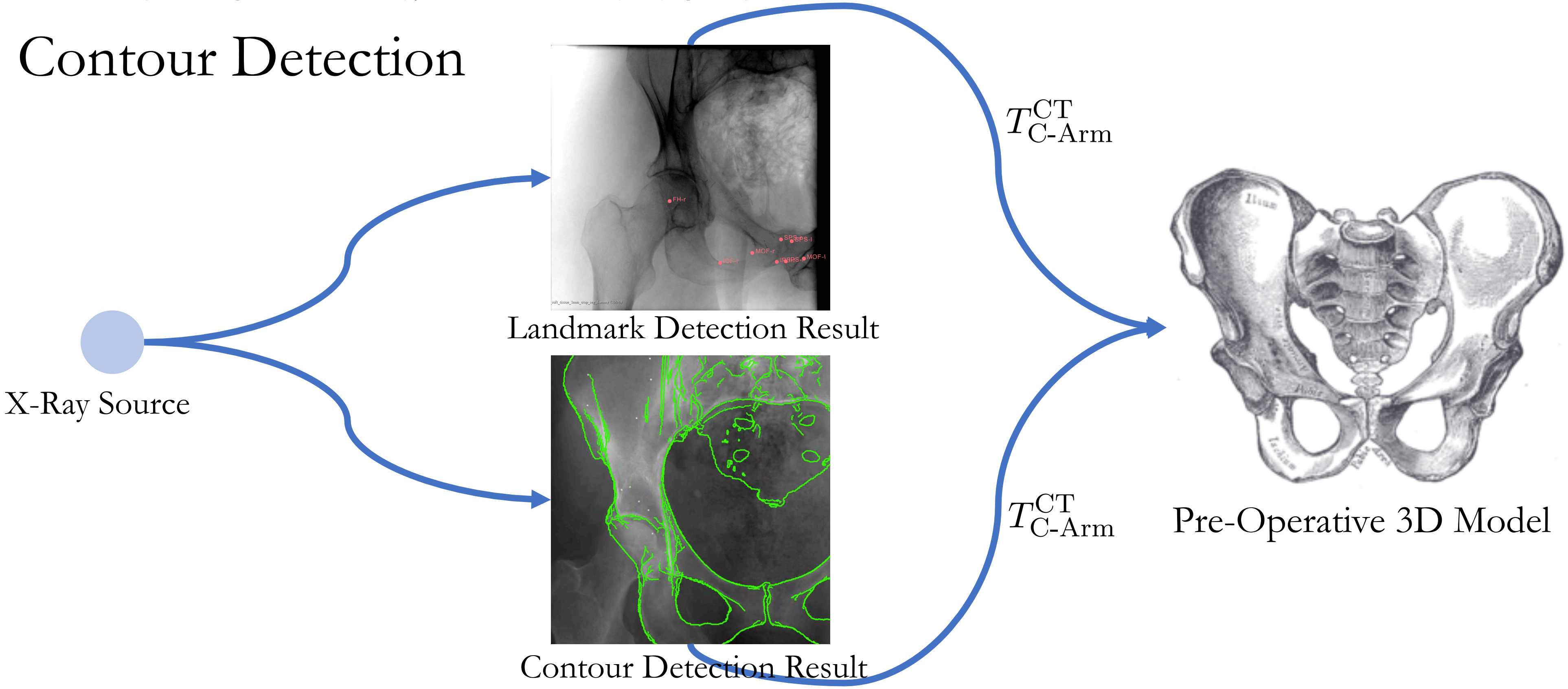
Mentor: Robert Grupp, Professor Russell Taylor

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Motivation

- Feature Detection for 2D-3D Registration
  - Anatomical Landmark Detection
  - Contour Detection



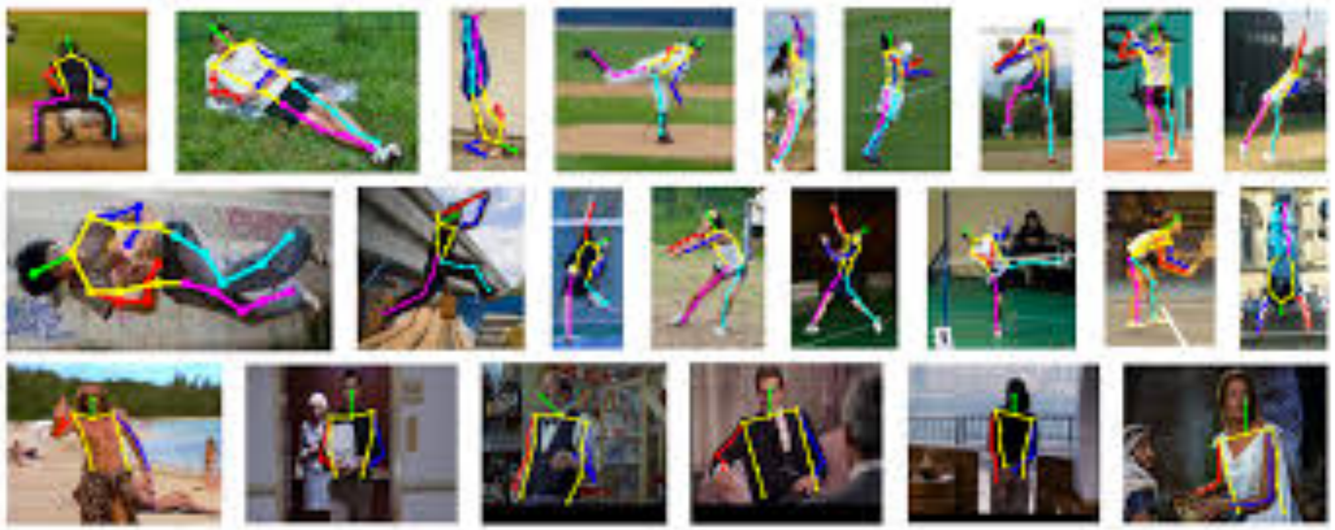
[1] Wikipedia contributors. (2018, March 17). Pelvis. In Wikipedia, The Free Encyclopedia. Retrieved 17:09, April 9, 2018.

Papers Discussed Today

- Human Pose Detection
- Wei, Shih-En, et al. "Convolutional pose machines." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
- Edge Detection
- Xie, Saining, and Zhuowen Tu. "Holistically-nested edge detection." Proceedings of the IEEE international conference on computer vision. 2015.

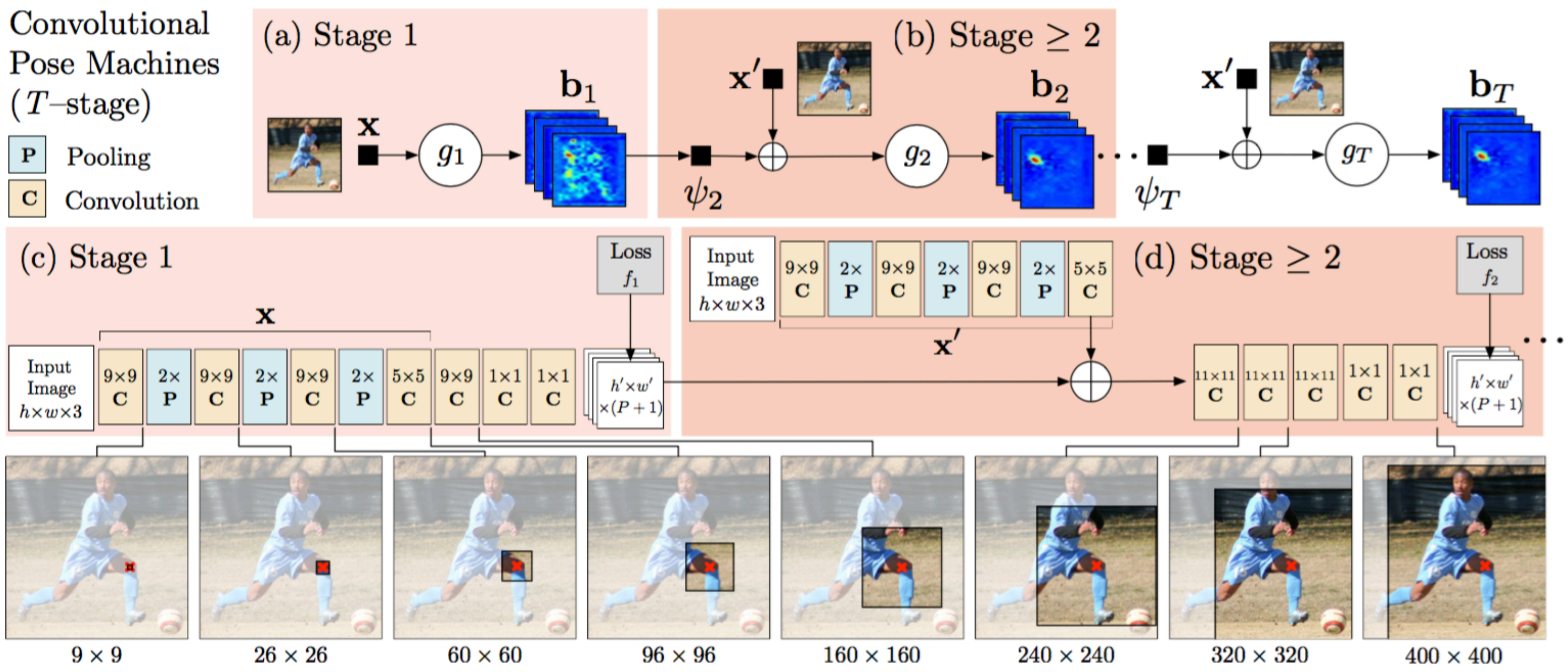
Convolutional Pose Machine

- Designed for human pose estimation task
- Sequential architecture composed of convolution layers to implicitly modeling long-range dependencies
- Perform belief map regression instead of pixel position
- Alleviate gradient vanishing by intermediate supervision



Network Architecture

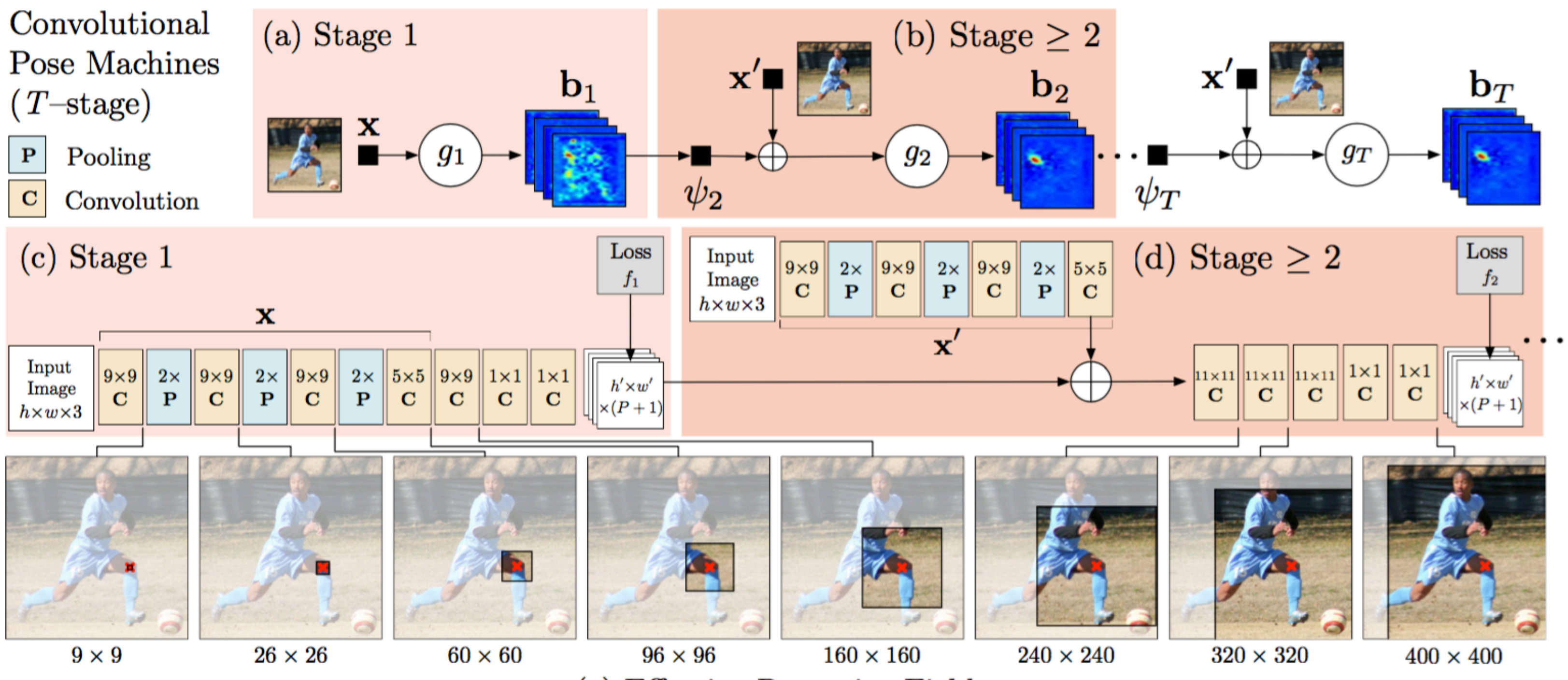
- Feature extractor at each stage (x)
  - Weight sharing among stages
  - Representation for the original image for following classification use



(e) Effective Receptive Field

Network Architecture

- Classifier at each stage (gt)
  - Takes previous stage output and feature extracted as input
  - Receptive field grows larger with deeper network



(e) Effective Receptive Field

Results and Discussions

- State of the art performance on single human pose estimation task.
- Weight sharing for feature extractor
- Inexplicit long range dependencies modeling
- Intermediate supervision (MSE loss for each stage)

$$F = \sum_{t}^{T} \sum_{p}^{P+1} \sum_{z} ||b_t^p(z) - b_*^p(z)||_2^2$$

Results and Discussions

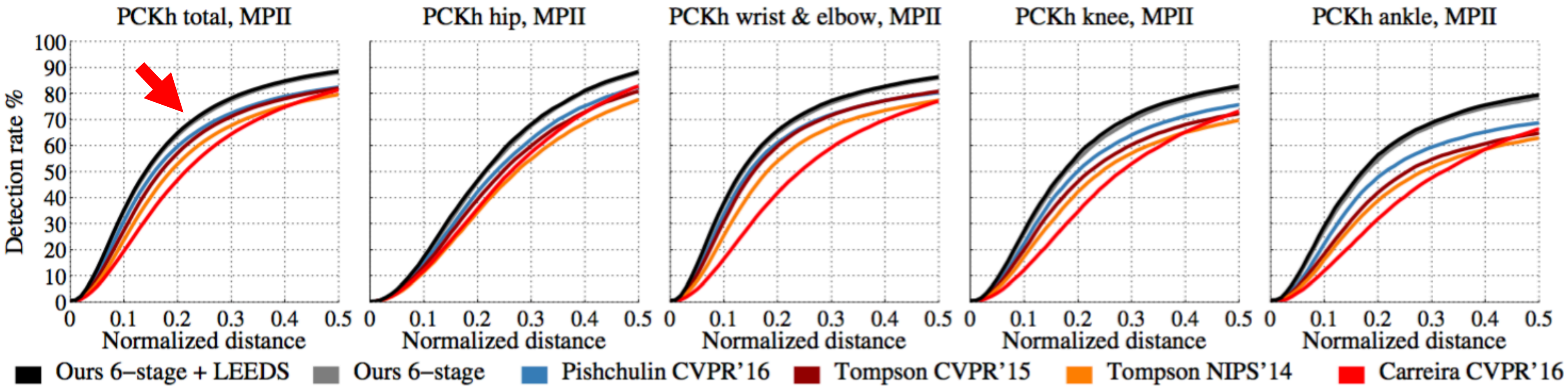


Figure 8: Quantitative results on the MPII dataset using the PCKh metric. We achieve state of the art performance and outperform significantly on difficult parts such as the ankle.

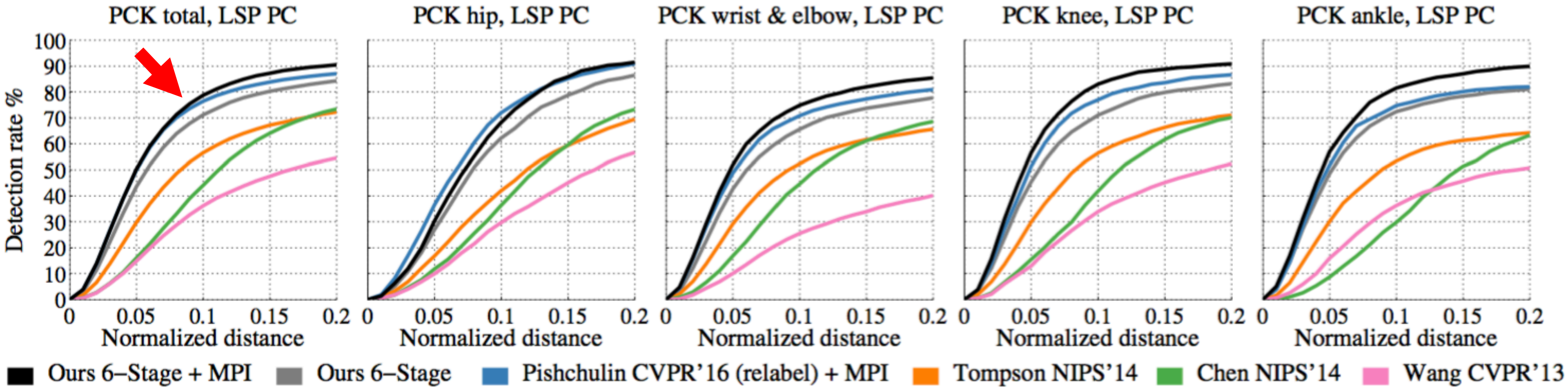


Figure 9: Quantitative results on the LSP dataset using the PCK metric. Our method again achieves state of the art performance and has a significant advantage on challenging parts.

Application

- Fixed number key point regression problem
- Belief map as uncertainly measure
  - Threshold poor predictions to make robust registration
  - Embed uncertainty into optimization objective function

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

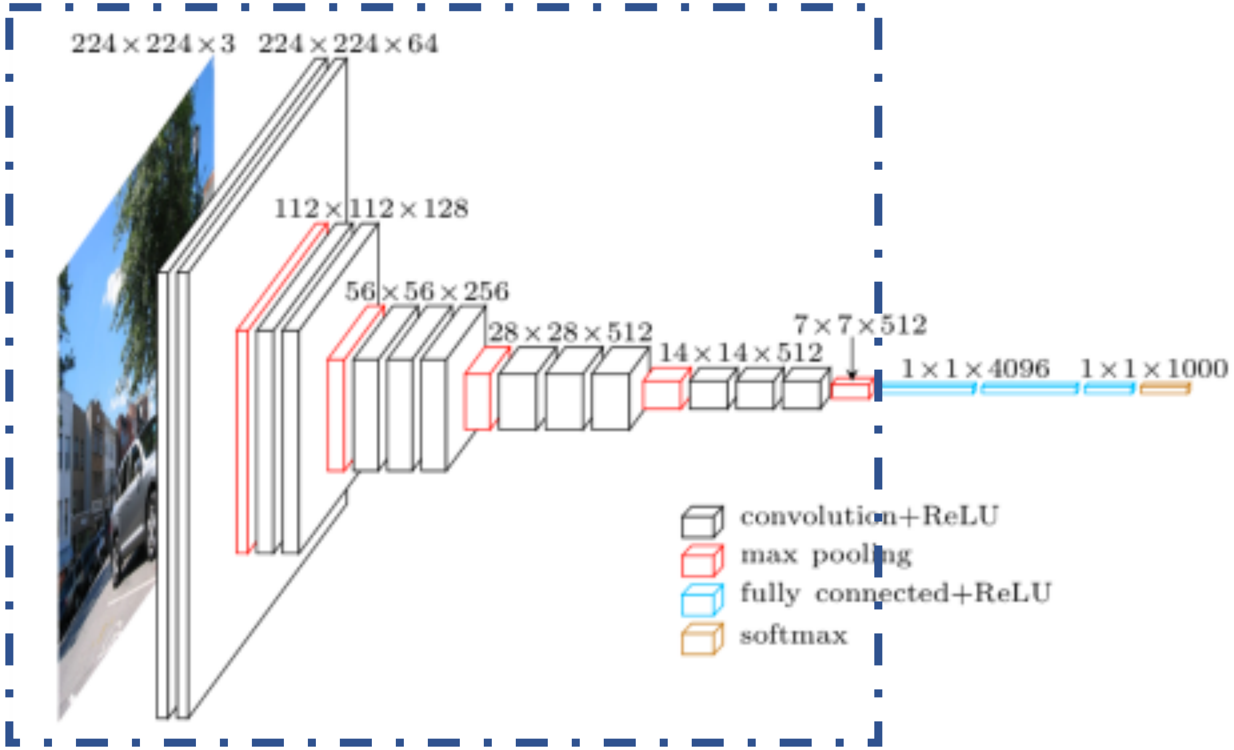
Holistically-Nested Edge Detection

- Holistic image training and prediction
- Multi scale feature learning through deep neural networks
- Completely differentiate framework
- Implicitly long range dependencies capture without graphical model

(a) original image (b) ground truth

Network Architecture

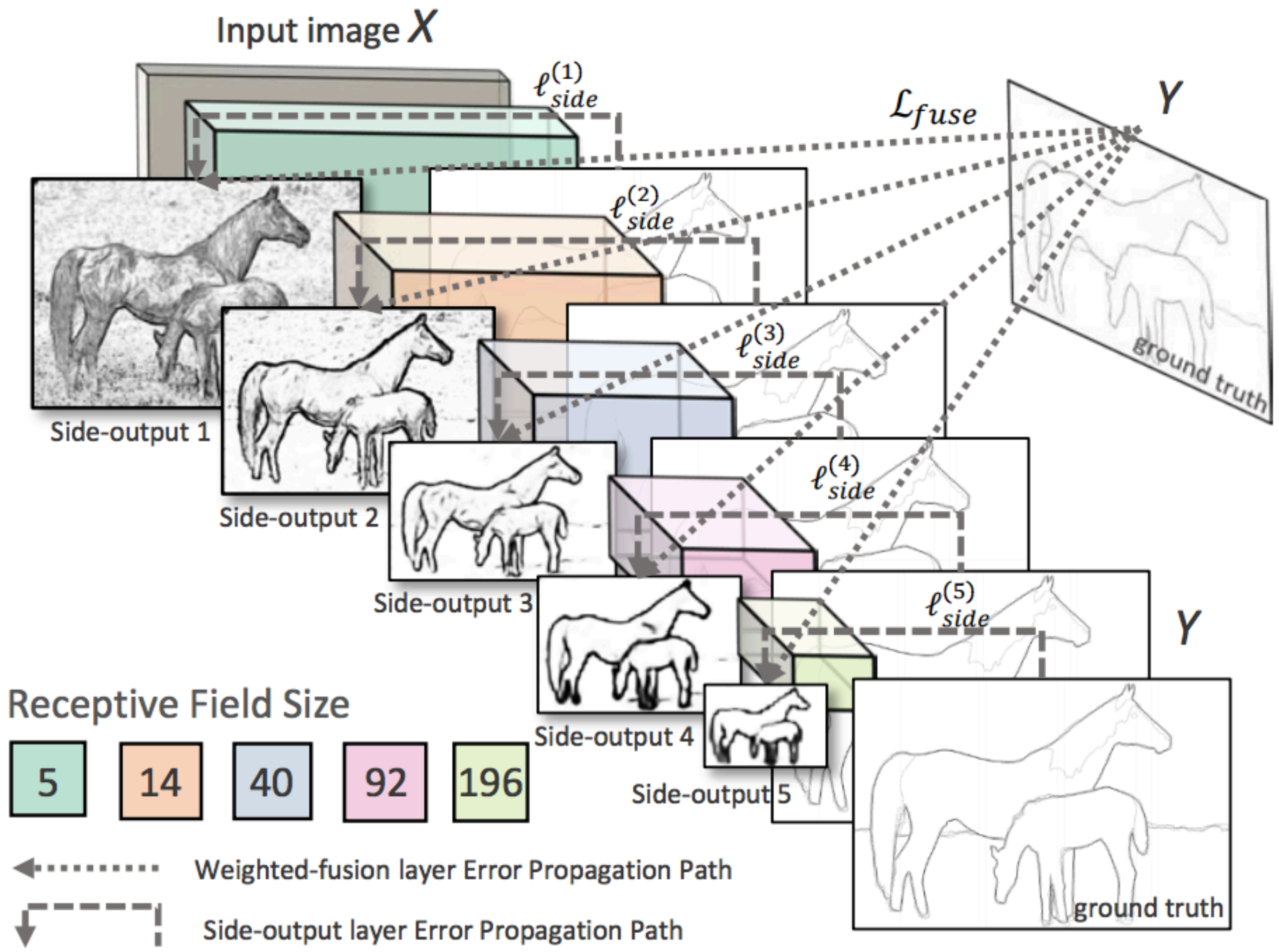
- Utilize VGG (convolutional parts) as basic block for feature extraction
  - With network going deeper, the features are extracted locally to globally



Network Architecture - Training

- Side output at each stage
  - Utilize conv layer as predictor here
- Loss at each stage
  - Weighted binary cross entropy loss

$$L = -\beta \sum_{Y_+} logP(y = 1|X; w)$$
$$-(1 - \beta) \textstyle\sum_{Y_-} logP(y = 0|X; w)$$



Network Architecture - Prediction

- Obtain side outputs and fused output from the network

$$Y^{fuse}, Y^{(1)}, \ldots, Y^{(M)} = HED(X.(W.w, h))$$

- Aggregate these predictions for final result

$$Y^{HED} = average(Y^{fuse}, Y^{(1)}, \ldots, Y^{(M)})$$

- Perform non maximum suppression for thinned edges result.

$$Y^{Final} = non\max suppression\,(Y^{HED})$$

Results

- State of the art performance
  - fixed contour threshold (ODS)
  - Per-image best threshold (OIS)
  - Average precision (AP)

Table 4. Results on BSDS500. ∗BSDS300 results,†GPU time

	ODS	OIS	AP	FPS
Human	.80	.80	-	-
Canny	.600	.640	.580	15
Felz-Hutt [9]	.610	.640	.560	10
BEL [5]	.660∗	-	-	1/10
gPb-owt-ucm [1]	.726	.757	.696	1/240
Sketch Tokens [24]	.727	.746	.780	1
SCG [31]	.739	.758	.773	1/280
SE-Var [6]	.746	.767	.803	2.5
OEF [13]	.749	.772	.817	-
DeepNets [21]	.738	.759	.758	1/5†
N4-Fields [10]	.753	.769	.784	1/6†
DeepEdge [2]	.753	.772	.807	$1/10^3$†
CSCNN [19]	.756	.775	.798	-
DeepContour [34]	.756	.773	.797	1/30†
HED (ours)	**.782**	**.804**	**.833**	2.5†, 1/12

Application

- Contour detection in X-Ray images
- Flexibility of basic blocks of HED
    - e.g. ResNet, Fully Convolutional Net, U-Net
- Predicted feature map has uncertainty measure about edges
    - Threshold weak edges
    - Embed into registration method

Conclusion

- Utilize deep neural network as feature extractor and classifier.
- Implicitly model long range dependencies by composing conv layers.
- Alleviate gradient vanishing by intermediate supervision.
- Uncertainty measure as extra information for registration.