Tracking of Orthopaedic Instruments in 3D Camera Views

Literature review by Jie Ying

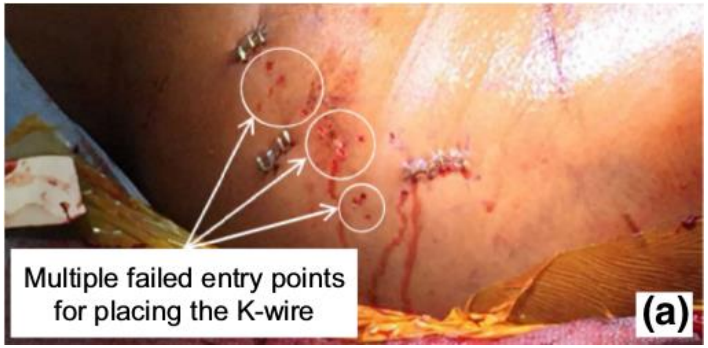
Group 3: Athira Jacob and Jie Ying Wu

Mentors: Dr. Bernhard Fuerst, Javad Fotouhi,
Mathias Unberath, Sing Chun Lee
Dr. Nassir Navab

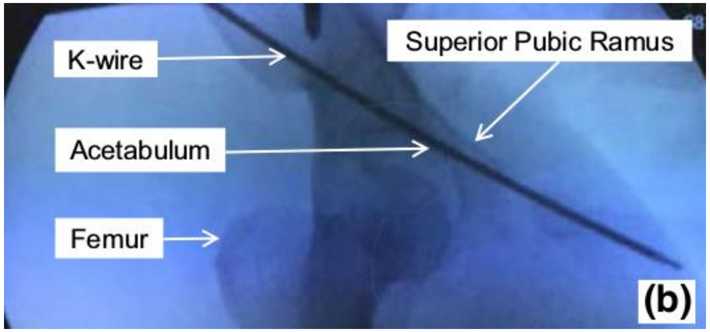
Technische Universität München + Johns Hopkins Whiting School of Engineering

Background

- K-wire insertion currently requires many X-rays
- Misplacement could damage important structures in the body
- Current tracking solutions are ineffective for K-wire
  - Traditional computer vision solutions fail
  - Trackers cannot be placed on it
- Propose to use convolutional neural network trained on RGB images



Multiple entry wounds



X-ray image of hip region in pelvic surgery

Images from Fischer, Marius, et al. "Preclinical usability study of multiple augmented reality concepts for K-wire placement." International Journal of Computer Assisted Radiology and Surgery 11.6 (2016): 1007-1014.

Tool tip prediction paper

Diotte, B., Fallavollita, P., Wang, L., Weidert, S., Thaller, P. H., Euler, E., & Navab, N., “Radiation-Free Drill Guidance in Interlocking of Intramedullary Nails,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, 2012, pp. 18–25.

Intramedullary nail placement in a phantom and the augmented drill position overlay

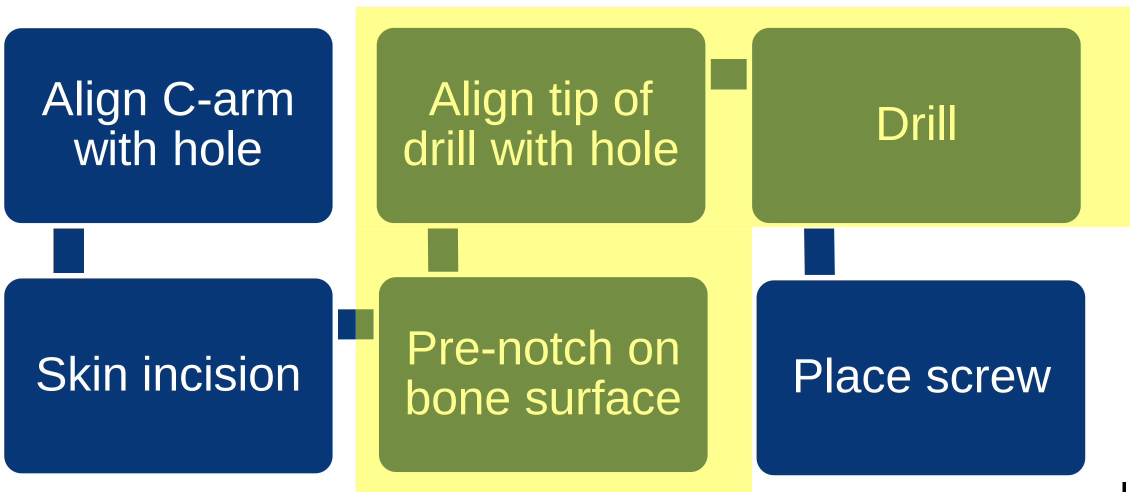
X-ray images of intramedullary nails in a fractured tibia

Image from *Diotte et al.* MICCAI 2012

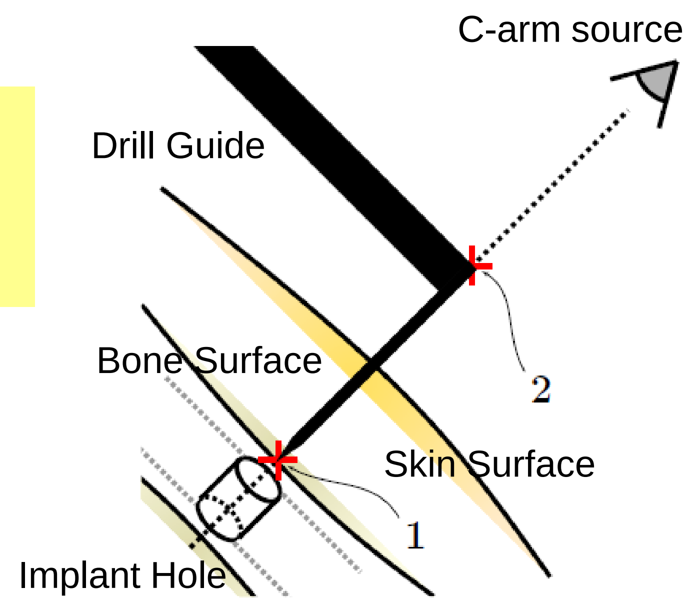
Images from Wikimedia Commons

Motivation

- Tibial fractures are common and treated with intramedullary nailing
- Can require 48 x-rays for one placement



Workflow of intramedullary nailing

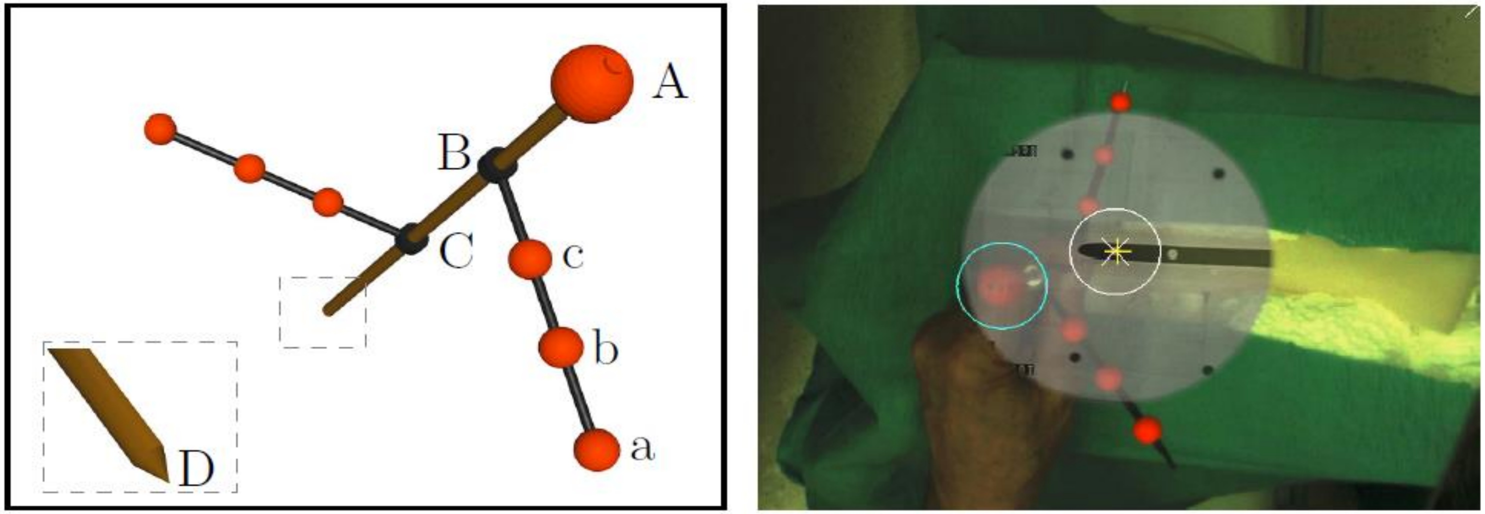


Desired alignment between drill and hole

- Propose radiation-free guidance based on optical marker tracking
- Augment projected tip position on the patient

Image from *Diotte et al.* MICCAI 2012

Method



Proposed optical marker to track the drill (left) and the augmentation that it provides (right). The tracked position of the top ball of the marker is circled in blue and the target drill position is circled in white. Estimated tip position is marked by the yellow cross.

$$d_{AD} = \frac{(S \cdot d_{AB}) - d_{AC}}{S - 1} \quad \text{where} \quad S = \frac{crossratio \cdot d_{AC}}{d_{AB}}$$

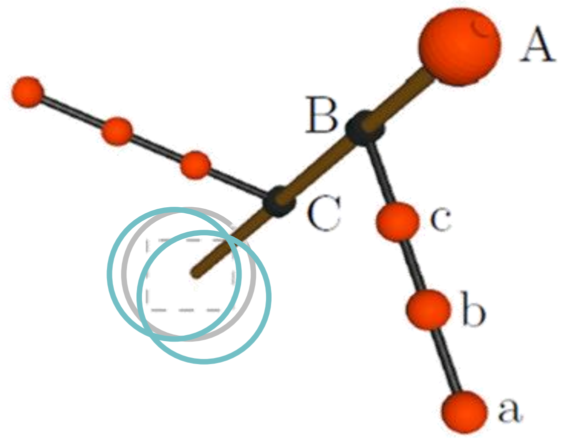
$$crossratio = \frac{AB \cdot CD}{AC \cdot BD}$$

Images from *Diotte et al.* MICCAI 2012

Experimental setup and results

- Print circle 5mm (size of nail hole)
- Fix the drill tip to the center of the circle
- Compare with estimated circle
- 200 trials over 30° cone angle rotation

- Mean error: 1.72 ± 0.7 mm
- 57% of the samples were below the mean
- 98% below 4mm



Marker image from *Diotte et al.* MICCAI 2012

Phantom experiment setup and results

- Phantom setup: dry bone fixed to a box with a 10mm Titanium Femoral Nail
- Fluoroscopic pre-drill image and one x-ray to confirm screw placement
- 3 surgeons – 2 experts and 1 resident – insert screw with marker guidance
- Success if they inserted the screw
- 93% success rate, 56/60 cases
- Failures attributed to resident and poor phantom setup

	Without guidance on real patient	With guidance on phantom
Num. of X-rays	48	2
Average time to completion	13.7 min	2 min

Image from *Diotte et al.* MICCAI 2012

Limitations

- Requires a large fixture to the drill
- Requires line of sight, mostly solved by mounting the camera on the C-arm
- Simpler problem here, only one correct placement
- More detailed algorithms needed
  - No discussion on registration algorithms
  - May contribute to errors in notching
- More detailed results would be nice
  - How many procedures were performed by experts vs. resident
  - How do timing vary between groups
  - What is the distribution of procedure time

Deep Learning Paper – Residual Learning

K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Proceedings of CVPR, 2016, pp. 770–778.

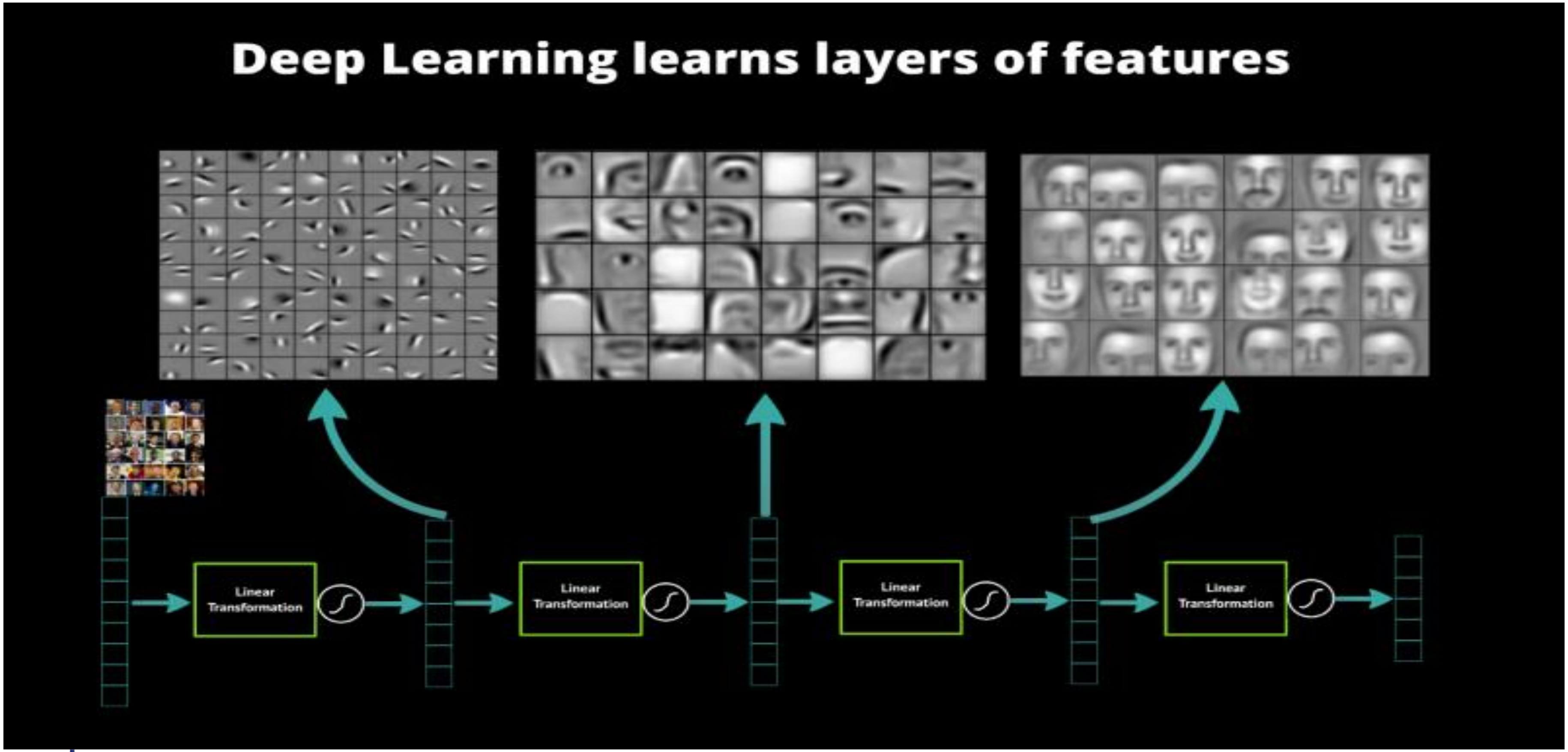
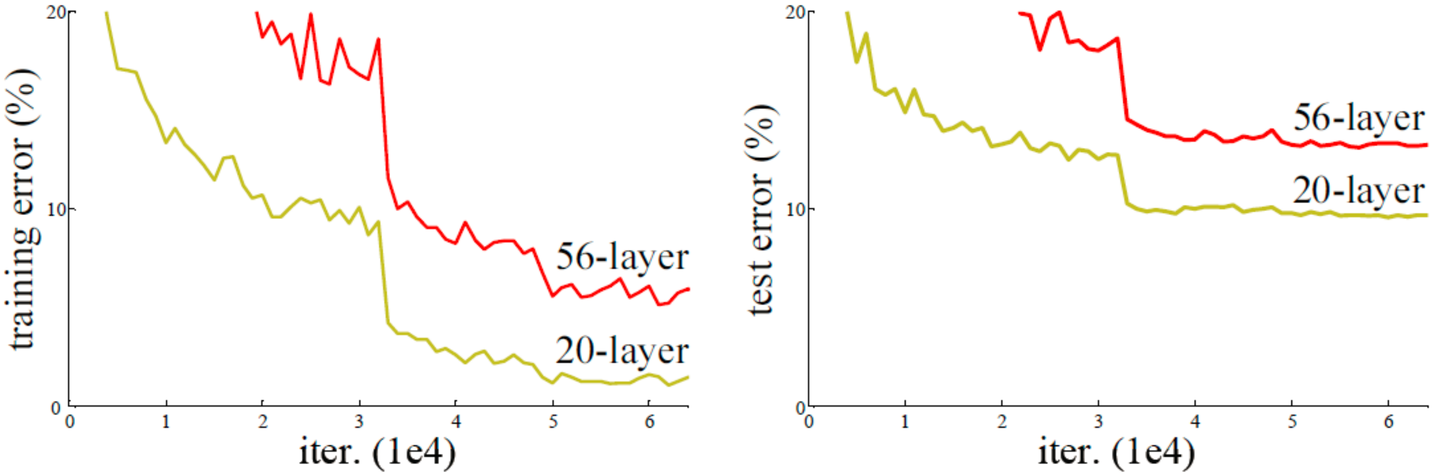


Image from D. Akagi, “A Primer on Deep Learning,” *DataRobot*, 19-May-2014. [Online]. Available: https://www.datarobot.com/blog/a-primer-on-deep-learning/.

Training a deep net

- Weight layers are expensive to learn
- Deeper nets have shown better performance but then degrades

- **Inspiration**: deepen a shallow network by padding with identity mapping
- Current solver cannot find do as well as or better than this construction



Training (left) and test (right) errors on CIFAR-10[1] with 20 and 56 layer "plain" (non-residual) networks. Even on training data, the deeper network has higher error.

[1] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images ," 2009.
Image from K. He, X. Zhang, S. Ren, and J. Sun, CVPR 2016.

Simplify the task

Suppose we have a filter

$$\mathcal{H}(x)$$

where x is the input to a layer

Instead of learning $\mathcal{H}(x)$, learn

$$\mathcal{F}(x) := \mathcal{H}(x) - x$$
$$\mathcal{H}(x) = \mathcal{F}(x) + x$$

Easy to learn identity mapping

$$\mathcal{F}(x) \rightarrow 0$$

No additional weights from the skip-ahead path since it is just identity

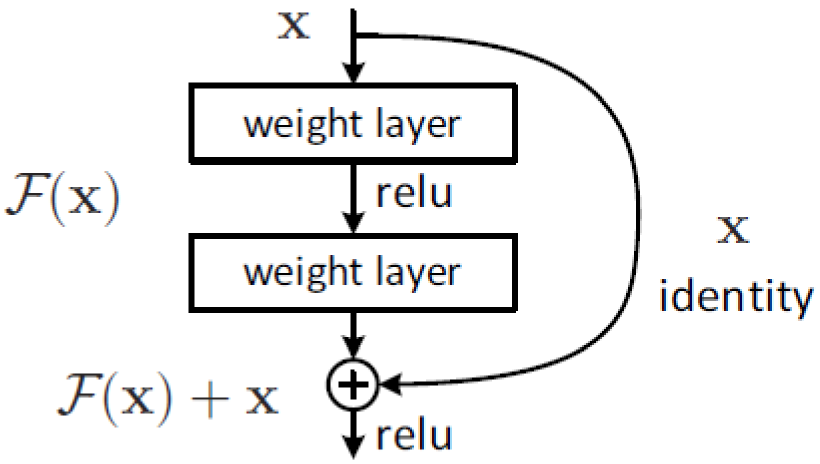
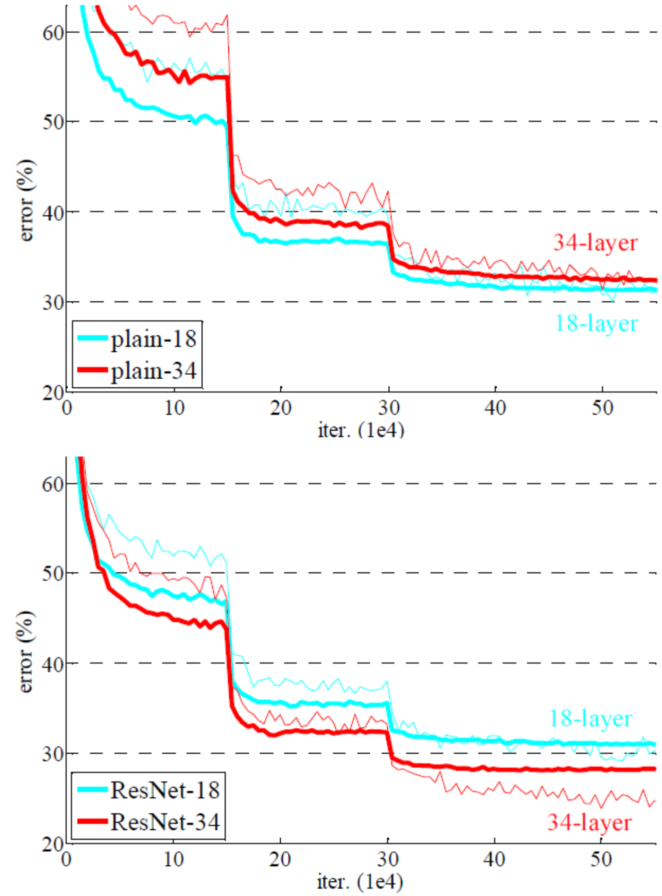


Image from K. He, X. Zhang, S. Ren, and J. Sun, CVPR 2016.

Results

Deeper residual nets do better

method	top-1 err.	top-5 err.
VGG [40] (ILSVRC'14)	-	8.43†
GoogLeNet [43] (ILSVRC'14)	-	7.89
VGG [40] (v5)	24.4	7.1
PReLU-net [12]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

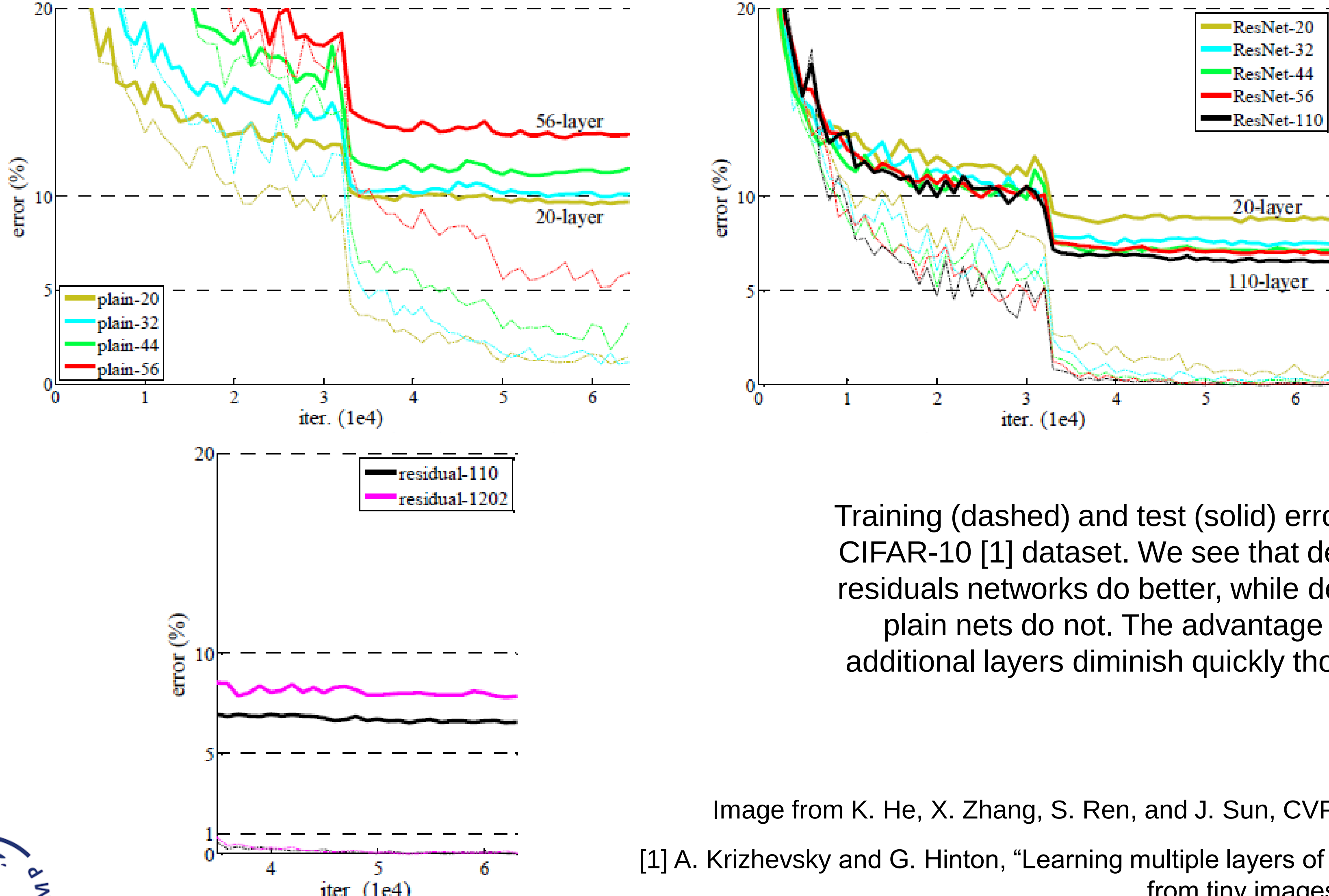


Comparing plain (top) and residual (bottom) networks of different depths trained and tested on ImageNet[1] . Thin lines show training error and bolded lines show test error.

[1] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int J Comput Vis*, Dec. 2015.

Image from K. He, X. Zhang, S. Ren, and J. Sun, CVPR 2016.

Results



Training (dashed) and test (solid) errors on CIFAR-10 [1] dataset. We see that deeper residuals networks do better, while deeper plain nets do not. The advantage of additional layers diminish quickly though.

Image from K. He, X. Zhang, S. Ren, and J. Sun, CVPR 2016.

[1] A. Krizhevsky and G. Hinton, “Learning multiple layers of features from tiny images,” 2009.

Limitations

- Output and input must be the same size – solve by projective mapping
- Unclear why this works, rarely need to learn identity mapping
  - Maybe just better initialization
  - Later paper claims the backpropagation works better
  - Solves vanishing gradients
- Object identification, not fully convoluted for pixel level labelling
- Claim faster convergence but no training time results
- Dropout often reduces training complexity – none used here