Automatic Segmentation of Chest Radiographs with Deep Learning

Greg Holste[1], Ryan Sullivan[2], Nick Nagy[3], Mike Bindschadler[3], Adam Alessio[4]
[1]Kenyon College, [2]Purdue University, [3]University of Washington, [4]Michigan State University

BACKGROUND

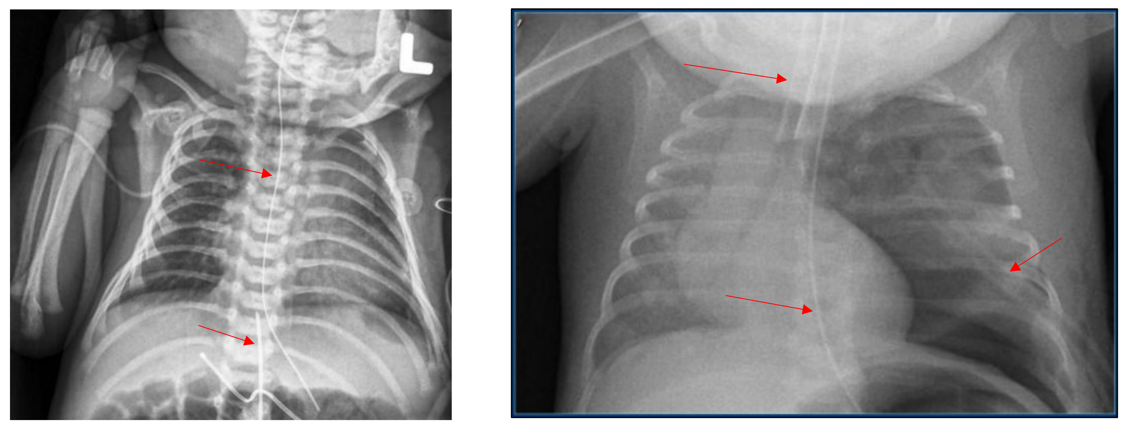

Figure 1. Two typical pediatric radiographs with lines marked by red arrows.

- Critically ill patients need catheters & tubes ("lines") to sustain life
  - monitoring their placement is time-consuming & labor-intensive
- Goal: automate with deep learning
  1. **segment chest into regions**
  2. find & classify lines
  3. determine if in correct location

METHODS

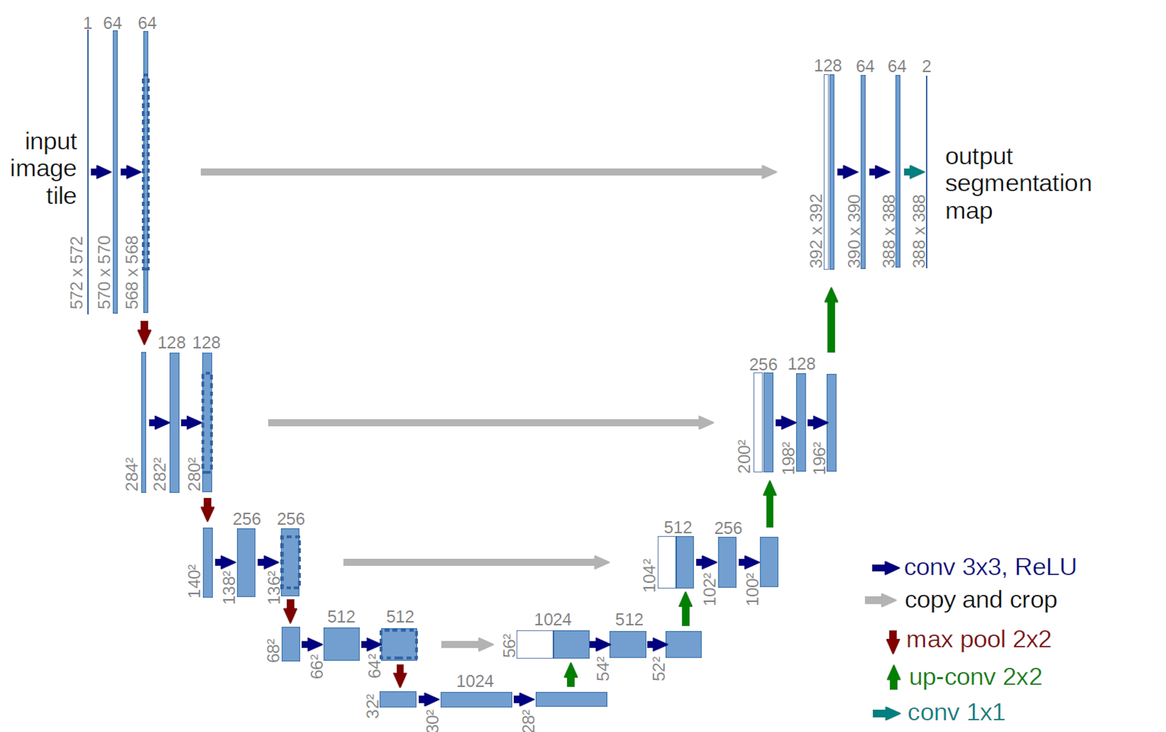

Figure 2. U-Net model architecture (Ronneberger *et al.* 2015). Fully convolutional network with a contracting (left) and expanding (right) path; the contracting path extracts features, while the expanding path up-samples those feature maps to the original input size and retains earlier information through skip connections from the contracting path.

- Trained U-Net on 300+ labeled radiographs
  - tested loss functions, weighting schemes, & data augmentation

Categorical cross-entropy: Generalized Dice loss:

$$-\sum_{n}\sum_{c} w_n(y_{n,c}\, log(p_{n,c})) \quad \text{versus} \quad 1-2\sum_{n}\sum_{c}\frac{w_n(y_{n,c}p_{n,c})}{w_n(y_{n,c}+p_{n,c})}$$

- n = pixel, c = class
- $y_{n,c}$ = indicator as to whether pixel n belongs to class c
- $p_{n,c}$ = probability that pixel n belongs to class c
- w_n = pre-computed "weight" of pixel n
  - 1 if unweighted

Deep **neural networks** can **automatically** find **regions** of the **chest** in pediatric **radiographs**.

PREDICTIONS

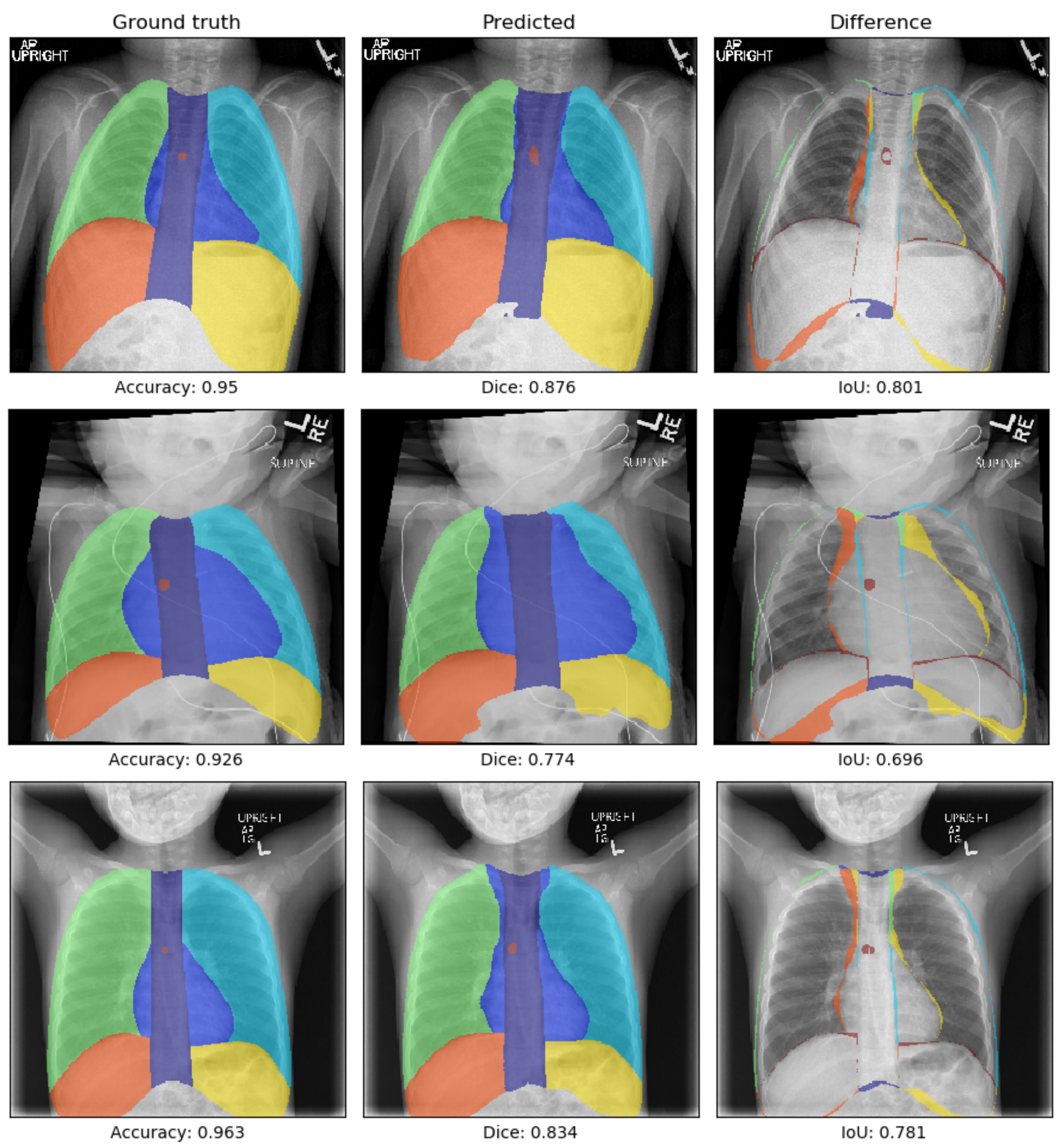

Figure 3. Best-performing model's predictions for three radiographs that it has not been trained on. Ground truth segmentation (left), predicted segmentation (middle) after thresholding probabilities, difference between predicted and true segmentations (right). Pixel accuracy, Dice coefficient, and intersection over union appear as axis labels on the left, middle, and right plots respectively.

Take a picture to see more predictions

RESULTS

- Categorical cross-entropy loss + data augmentation provided best results
  - pixel weights gave small improvement
  - $\overline{\text{Dice}}$: 0.832, $\overline{\text{Accuracy}}$: 0.938

Loss	Dice
w2-cce*	0.832
w3-cce*	0.83
w1-cce*	0.829
cce + gdl*	0.828
cce*	0.828
w2-cce	0.825
cce + gdl	0.821
w3-cce	0.82
cce	0.815
w1-cce	0.814
w3-gdl	0.788
w1-gdl	0.787
gdl	0.784
w2-gdl	0.783

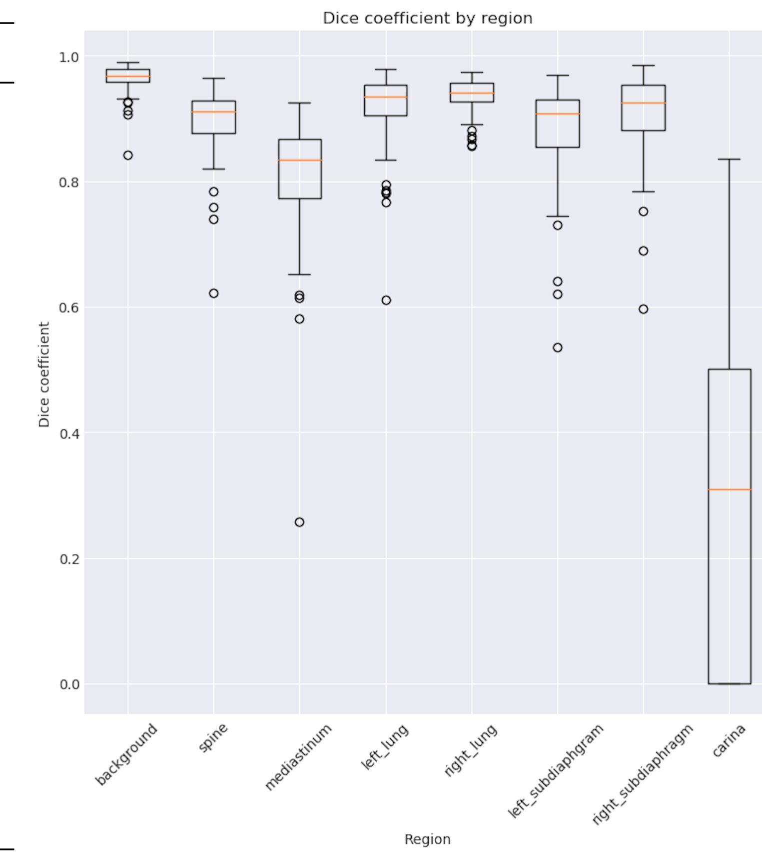

Figure 4. Mean Dice coefficient on test set (n=70) by loss function (left) and box-and-whisker plot of Dice coefficients by class for the best model (right). "*" = with data augmentation, "cce" = categorical cross-entropy, "gdl" = generalized Dice loss, "w<no.>" = with pre-computed pixel weights from method <no.>.

FUTURE WORK

- Train on larger-resolution images
- Obtain more, higher-quality labels
  - or… more aggressive augmentation
- Ensure predictions are biologically sound
  - guarantee single instance of each class
  - prohibit disconnected regions
- Combine with line detector

REFERENCES

- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention.* Springer, Cham, 2015.
- Crum, William R., Oscar Camara, and Derek LG Hill. "Generalized overlap measures for evaluation and validation in medical image analysis." *IEEE transactions on medical imaging* 25.11 (2006): 1451-1461.

ACKNOWLEDGMENTS

We acknowledge support from the iCER ACRES REU at MSU, which is funded by the National Science Foundation through grant 1560168.

Advanced Computational Research Experience for Students