CROSS MODALITY SYNTHESIS (MRI TO CT)

Yu Zhao, zhaoyu.hust@gmail.com

Background

- Radiotherapy treatment planning:
  - Target delineation, treatment response assessment, adaptive therapies, etc.
- MR images are necessary: superior soft-tissue contrast, etc. Missing: electron density info, provided by CT images → CT-MRI coregistration: 1. misalignment; 2. workflow, patients and costs pressure.
- MRI-only simulation workflow: providing both info on tumor volume and location and electron density information.

Background

- CT-MR registration VS MR-only simulation

CT-MR registration workflow

CT → MRI → Registration → Delineation → Dose Plan

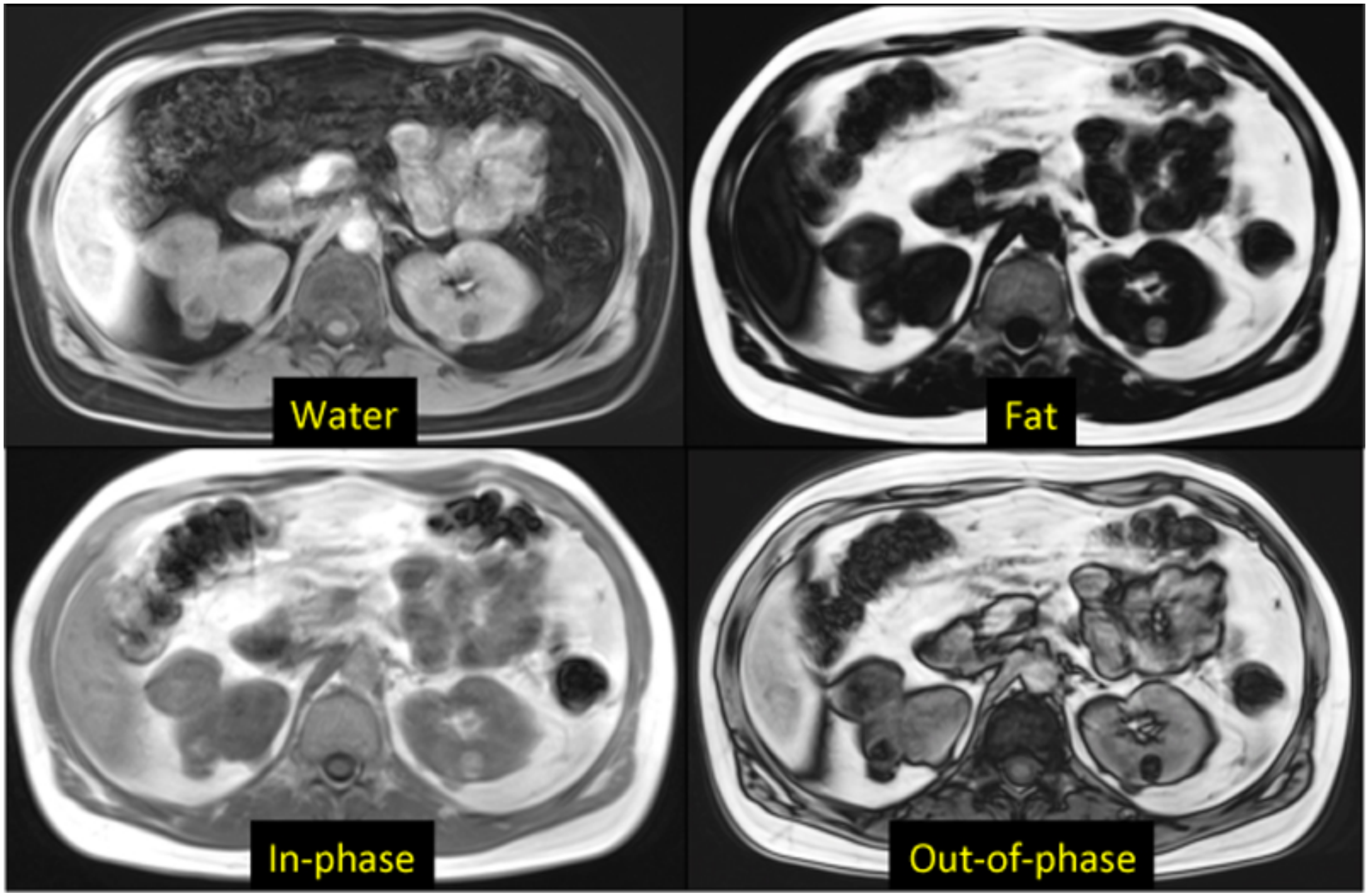
MR-only simulation workflow

MRI+MRCAT → Delineation → Dose Plan

Background

- MRI images: in-phase (IP), out-of-phase (OOP): sequences correspond to paired MRI gradient echo (GRE) sequences obtained with the same repetition time (TR) but with two different echo time (TE)values.
- The main application of the IP-OOP sequences is to identify pathological (microsopic) fat content of tissues in the abdomen by showing signal intensities drop on the OOP images compared to the IP images.

Background



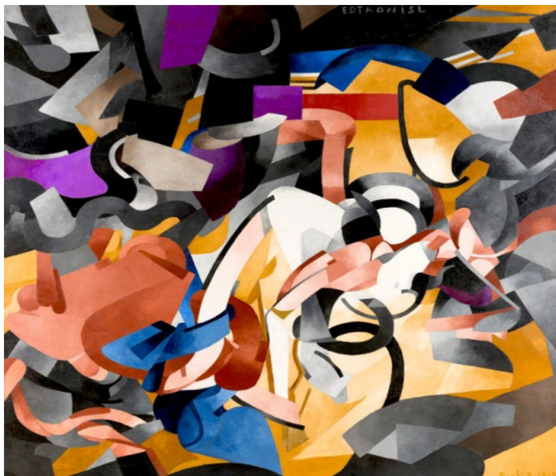
fat only = in-phase - opposed phase
water only = in-phase + opposed phase

Methods

- 1. Style Transfer
- 2. Image Analogy
- 3. Deep Learning Nets

1. Style Transfer

Style



Content

Synthetic

1. Style Transfer

- **A Neural Algorithm of Artistic Style**

  Leon A. Gatys, Alexander S. Ecker, Matthias Bethge

- 1). A new way to visualize feature. Use gradient decent to update the image till it generate similar feature response.
- 2). content reconstruction
- 3). style reconstruction

1. Style Transfer

$\mathcal{L}_{\text{content}}\Big($  ,  $\Big) \approx 0$

A schematic of the content loss.

$\mathcal{L}_{\text{style}}\Big($  ,  $\Big) \approx 0$

A schematic of the style loss.

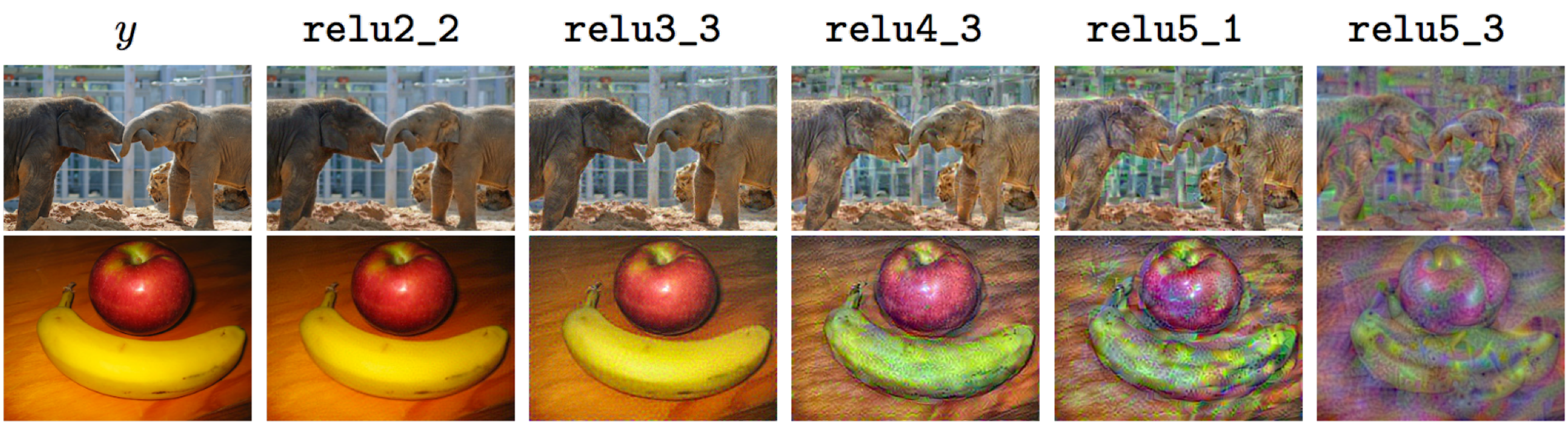
1. Style Transfer

- Feature extraction nets for losses

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

1. Style Transfer

Feature visualization



Content features at different level

Content loss

- The content loss is the (scaled, squared) Euclidean distance between feature representations of the content and combination images.
- draw the content feature from block2_conv2 ➔ structural detail

Style loss

Gram matrix

- The terms of this matrix are proportional to the covariances of corresponding sets of features, and thus captures information about which features tend to activate together. By only capturing these aggregate statistics across the image, they are blind to the specific arrangement of objects inside the image. This is what allows them to capture information about style independent of content.
- The style loss is then the (scaled, squared) Frobenius norm of the difference between the Gram matrices of the style and combination images.

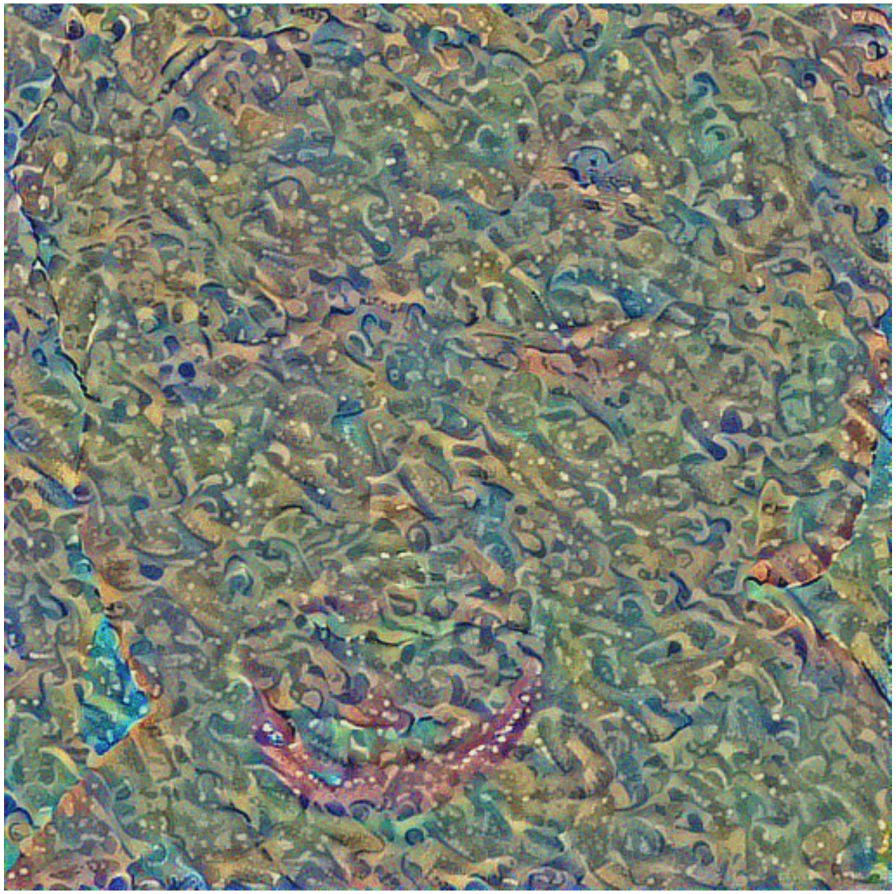
Total loss

- Total loss

= content loss + style loss + variation loss

(a regularization term that encourages spatial smoothness)

- Regression
- Gradient Decent



Reproduction results

Style Image

Content Images

2. Using Image Analogies:

- http://www.mrl.nyu.edu/publications/image-analogies/analogies-fullres.pdf

Figure 1 An image analogy. Our problem is to compute a new "analogous" image B' that relates to B in "the same way" as A' relates to A. Here, A, A', and B are inputs to our algorithm, and B' is the output. The full-size images are shown in Figures 10 and 11.

3. Deep Learning Nets

U-Net: Convolutional Networks for Biomedical Image Segmentation

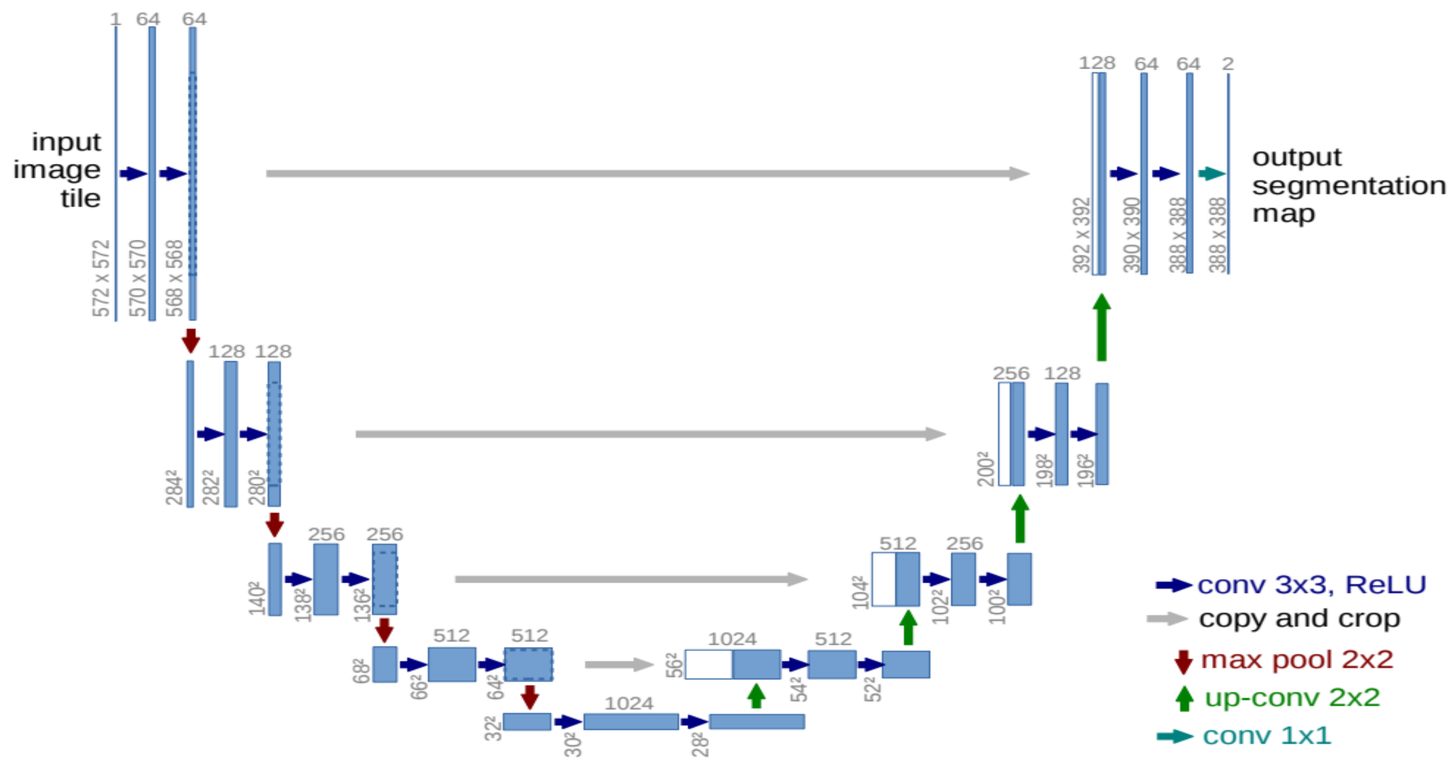


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

3. Deep Learning Nets

U-Net: Convolutional Networks for Biomedical Image Segmentation

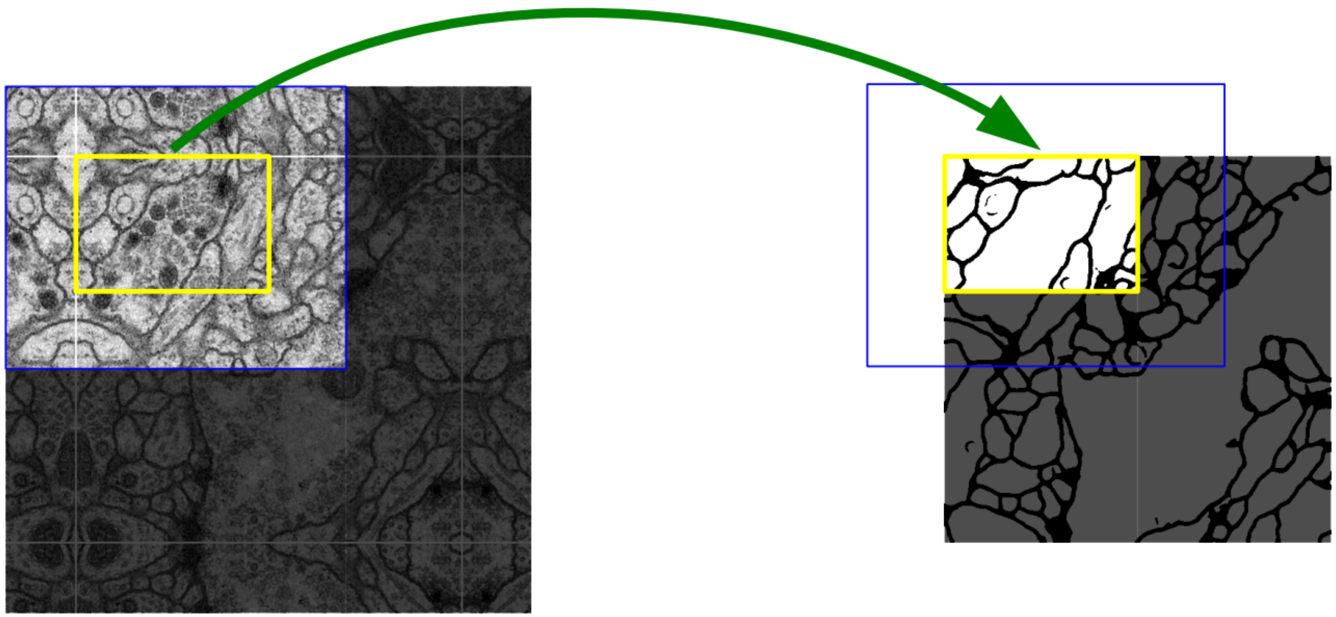


Fig. 2. Overlap-tile strategy for seamless segmentation of arbitrary large images (here segmentation of neuronal structures in EM stacks). Prediction of the segmentation in the yellow area, requires image data within the blue area as input. Missing input data is extrapolated by mirroring

3. Deep Learning Nets

U-Net: Convolutional Networks for Biomedical Image Segmentation

$$E = \sum_{\mathbf{x} \in \Omega} w(\mathbf{x}) \log(p_{\ell(\mathbf{x})}(\mathbf{x}))$$

pk(x) ≈ 1 for the k that has the maximum activation ak(x) and
pk(x) ≈ 0 for all other k.

3. Deep Learning Nets

U-Net: Convolutional Networks for Biomedical Image Segmentation

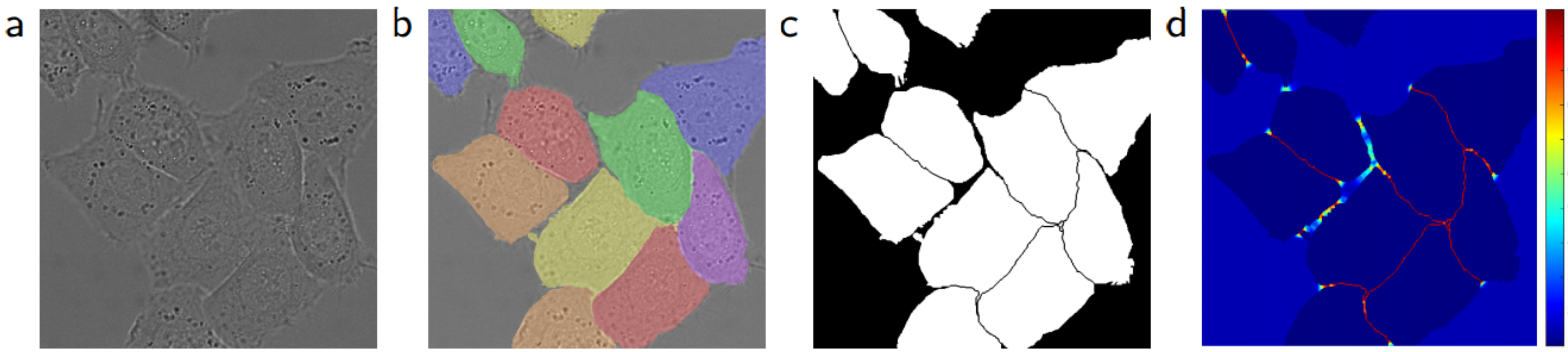


Fig. 3. HeLa cells on glass recorded with DIC (differential interference contrast) microscopy. (**a**) raw image. (**b**) overlay with ground truth segmentation. Different colors indicate different instances of the HeLa cells. (**c**) generated segmentation mask (white: foreground, black: background). (**d**) map with a pixel-wise loss weight to force the network to learn the border pixels.

3. Deep Learning Nets

U-Net: Convolutional Networks for Biomedical Image Segmentation

Data augmentation is essential to teach the network the desired invariance and robustness properties, when only few training samples are available. Especially random elastic deformations of the training samples seem to be the key concept to train a segmentation network with very few annotated images. We generate smooth deformations using random displacement vectors on a coarse 3 by 3 grid. The displacements are sampled from a Gaussian distribution with 10 pixels standard deviation.

3. Deep Learning Nets

U-Net: Convolutional Networks for Biomedical Image Segmentation

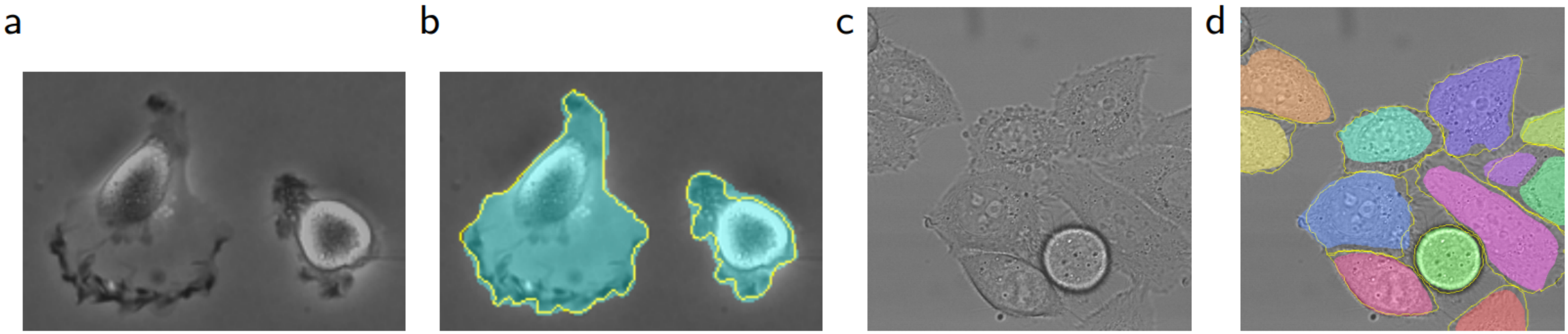


Fig. 4. Result on the ISBI cell tracking challenge. (**a**) part of an input image of the "PhC-U373" data set. (**b**) Segmentation result (cyan mask) with manual ground truth (yellow border) (**c**) input image of the "DIC-HeLa" data set. (**d**) Segmentation result (random colored masks) with manual ground truth (yellow border).

3. Deep Learning Nets

V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation

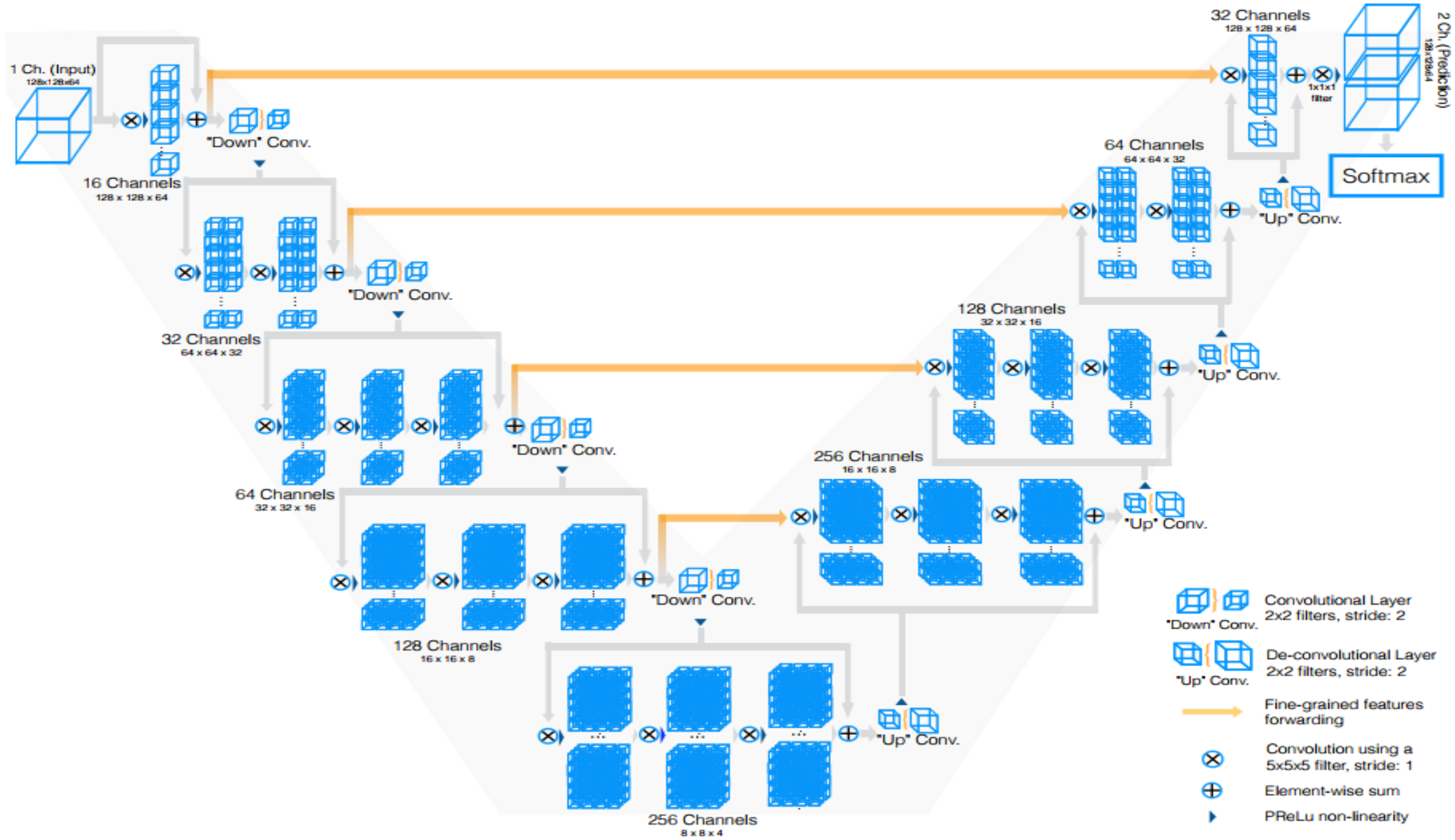


Fig. 2. Schematic representation of our network architecture. Our custom implementation of Caffe [5] processes 3D data by performing volumetric convolutions. Best viewed in electronic format.

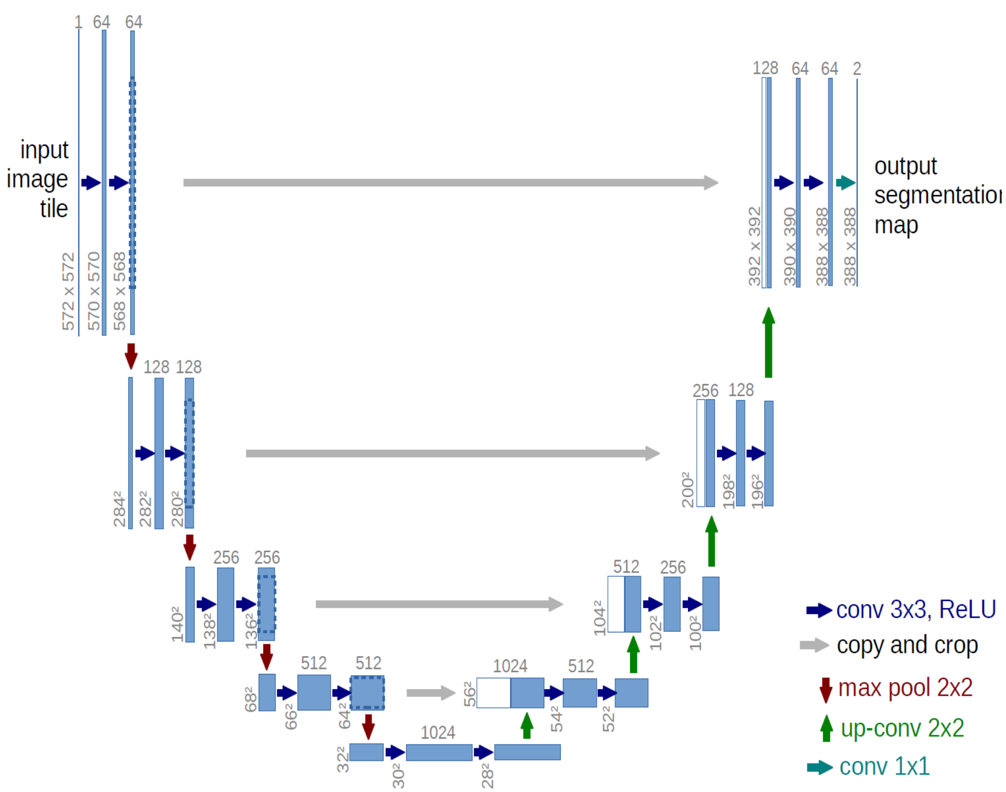
3. Deep Learning Nets

- Deep learning for MRI-CT synthesis for radiotherapy
  - perfectly registered CT-MR pairs
  - Memory issue for training of a standard 3D deep network
- Multi-view Deep Convolutional Neural Networks
  - Effective cost function for misalignments
  - Multi-view combination for memory efficiency without losing 3D context

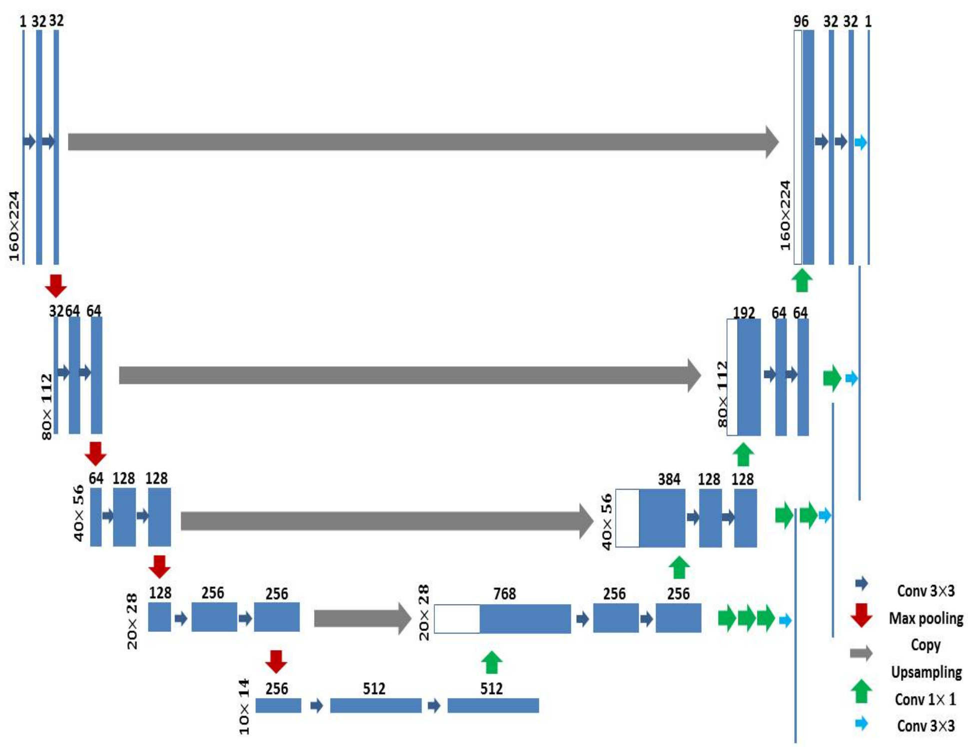
3. Deep Learning Nets

1). Unet based structure:

U-Net: Convolutional Networks for Biomedical Image Segmentation

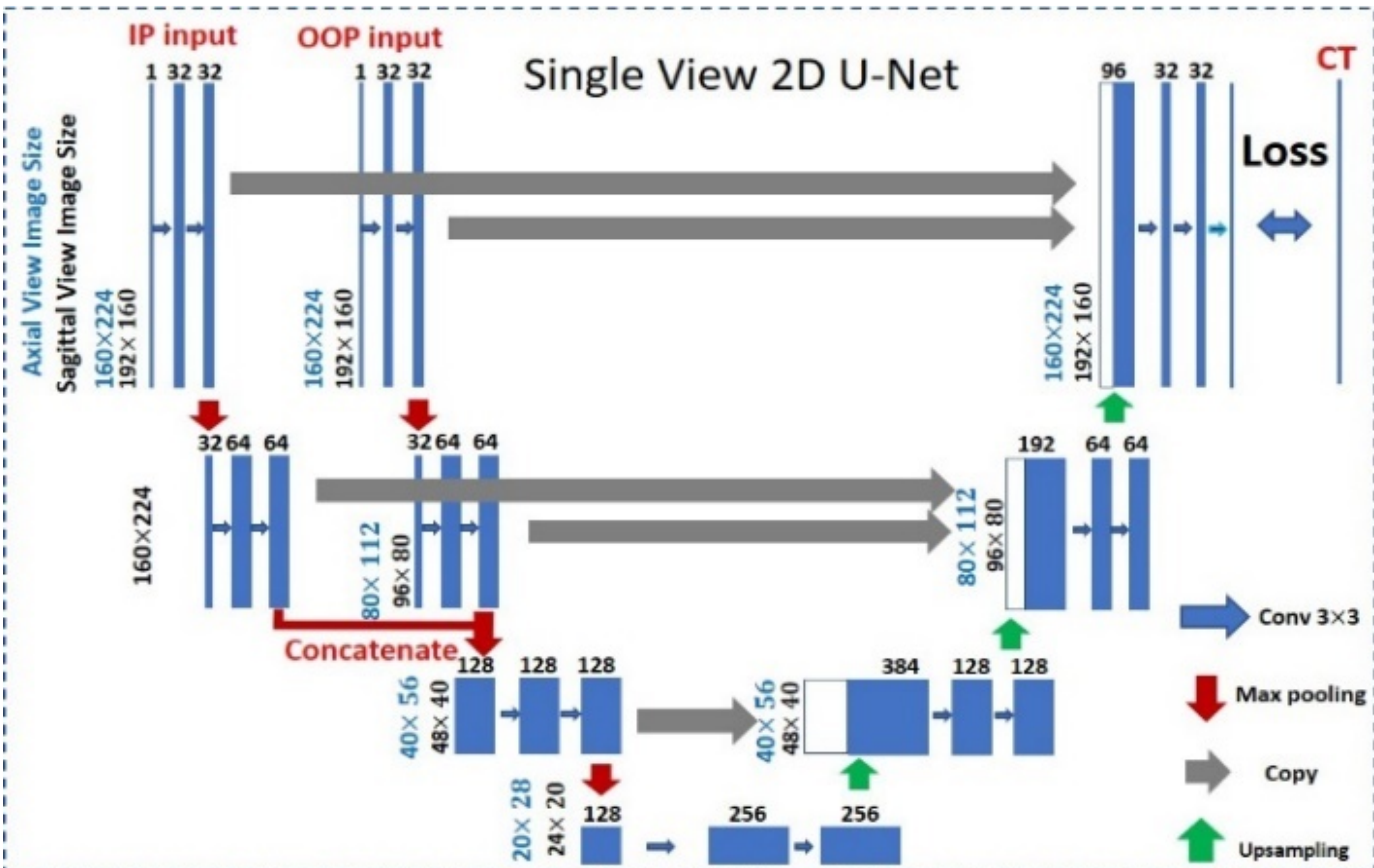


U-Net: Our modified version



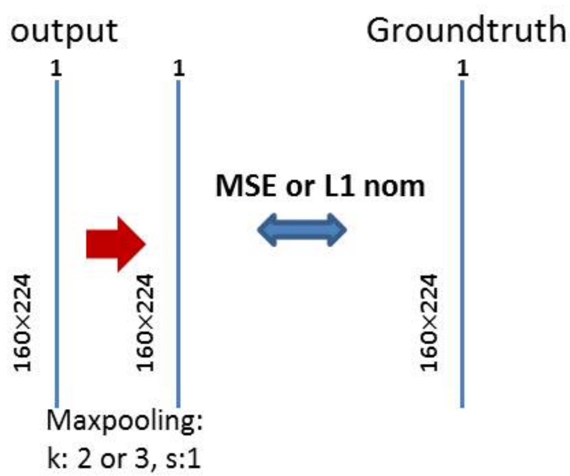
3. Deep Learning Nets

3. Deep Learning Nets



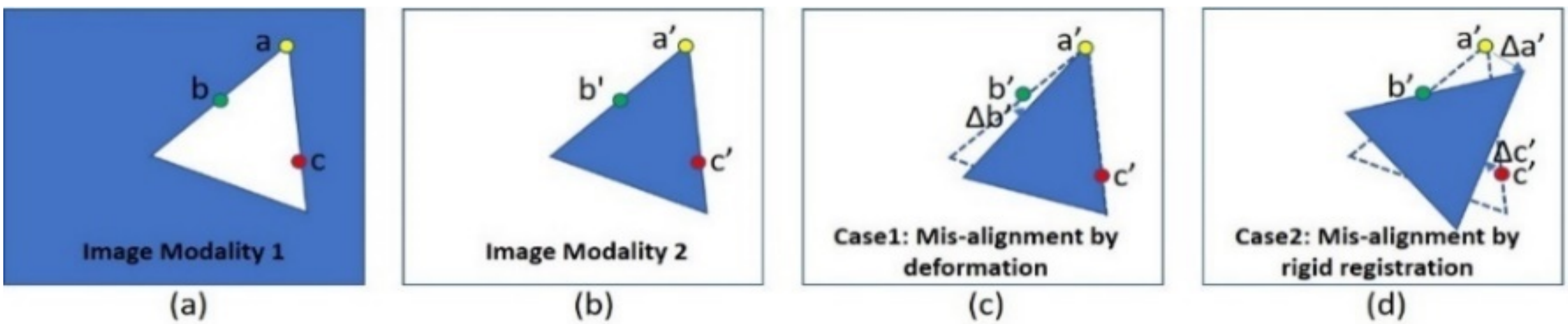
3. Deep Learning Nets

- Misalignment between Ground Truth and Inputs -- Max Pooled loss:
- 1). Maxpooling before calculating actual loss



3. Deep Learning Nets

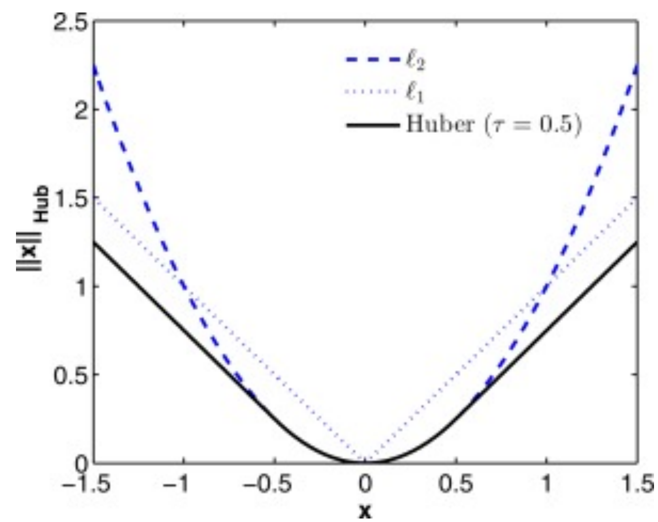
- Misalignment between Ground Truth and Inputs -- Max Pooled loss:



A schematic explanation of the impact of mis-registration to intensity transformation. (a) Image1 (Modality 1), (b) A perfectly registered Image2 (Modality 2), (c) Image 2 with rigid mis-alignment, (d) Image 2 with non-rigid mis-alignment. Triangles in (a)-(d) represent the same object. Dashed lines in (c) and (d) denote the locations of the perfectly registered Image2.

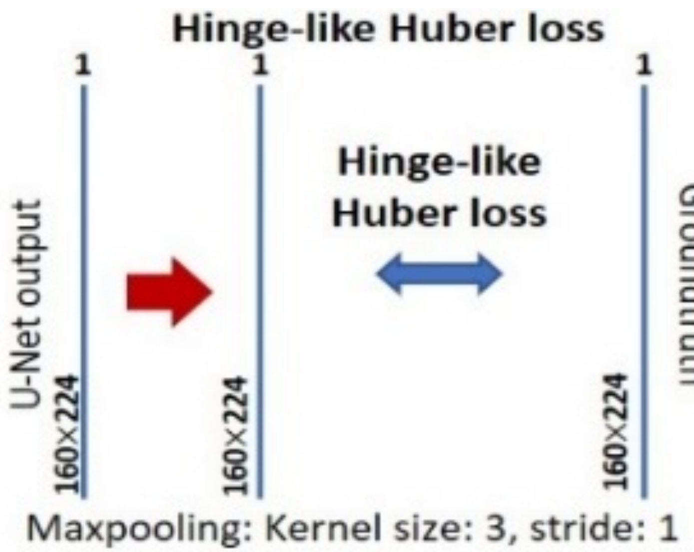
3. Deep Learning Nets

- Misalignment between Ground Truth and Inputs -- Max Pooled loss:
- 2). using L1 distance rather than L2 as L1 encourages less blurring (Image-to-Image Translation with Conditional Adversarial Networks, Phillip Isola, et al, 2016)

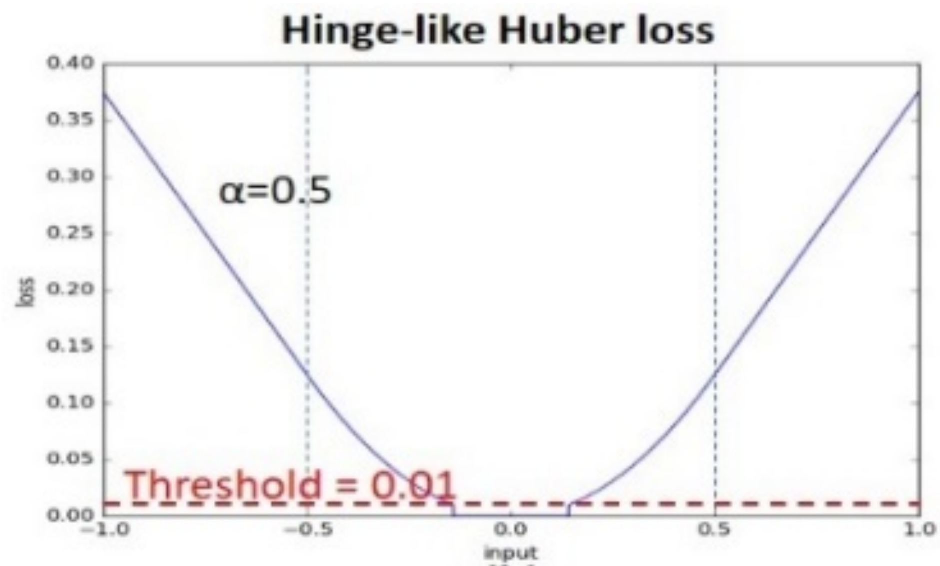


3. Deep Learning Nets

- Final loss: Maxpooling Hinge-like Huber loss



(a)

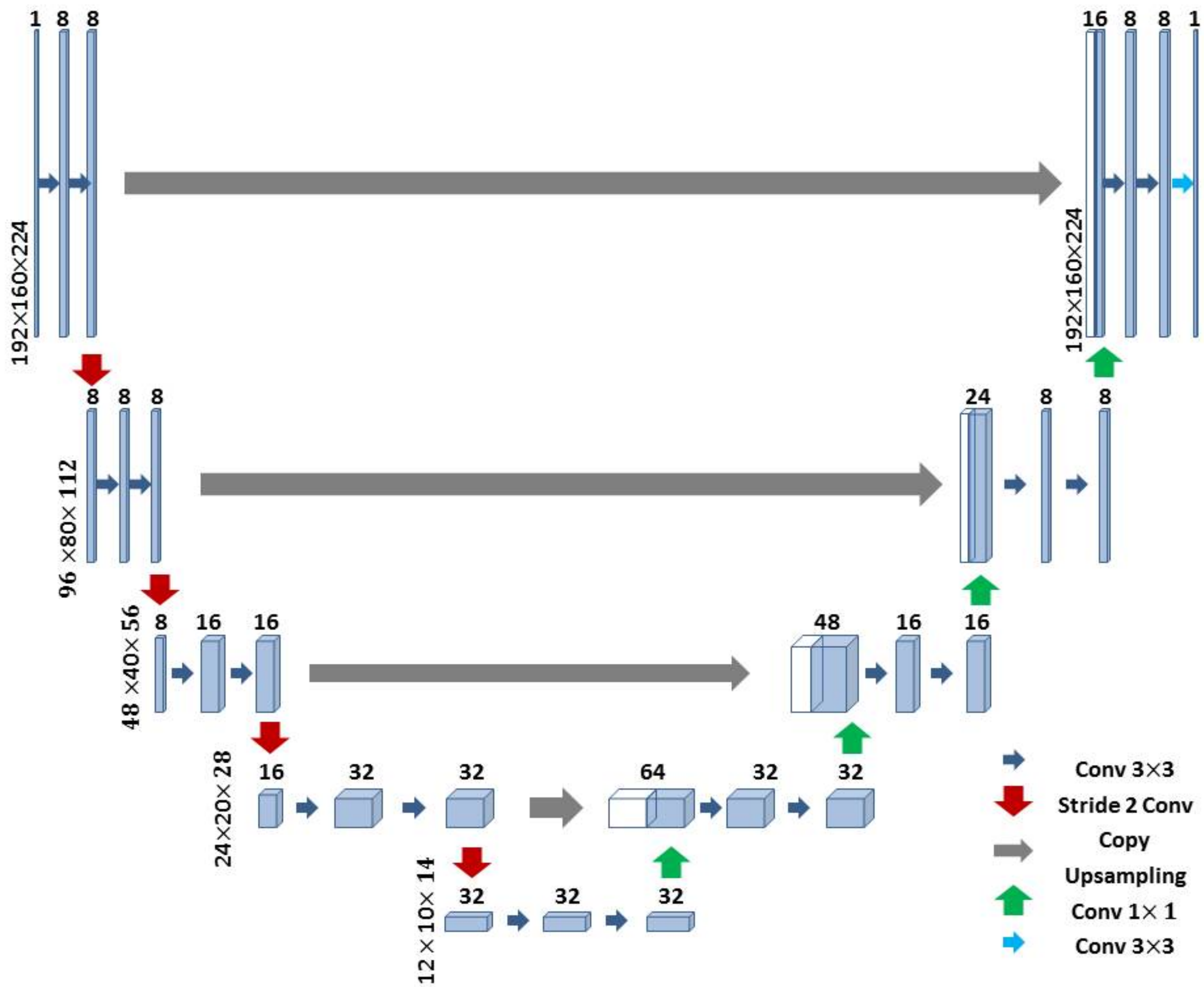


(b)

$$L(a) = \begin{cases} 0, & |a| < 0.01 \\ \frac{1}{2}a^2, 0.01 < |a| < \alpha \\ \alpha\left(|a| - \frac{1}{2}\alpha\right), otherwise \end{cases}$$

3. Deep Learning Nets

- 3D models (whole volume in)

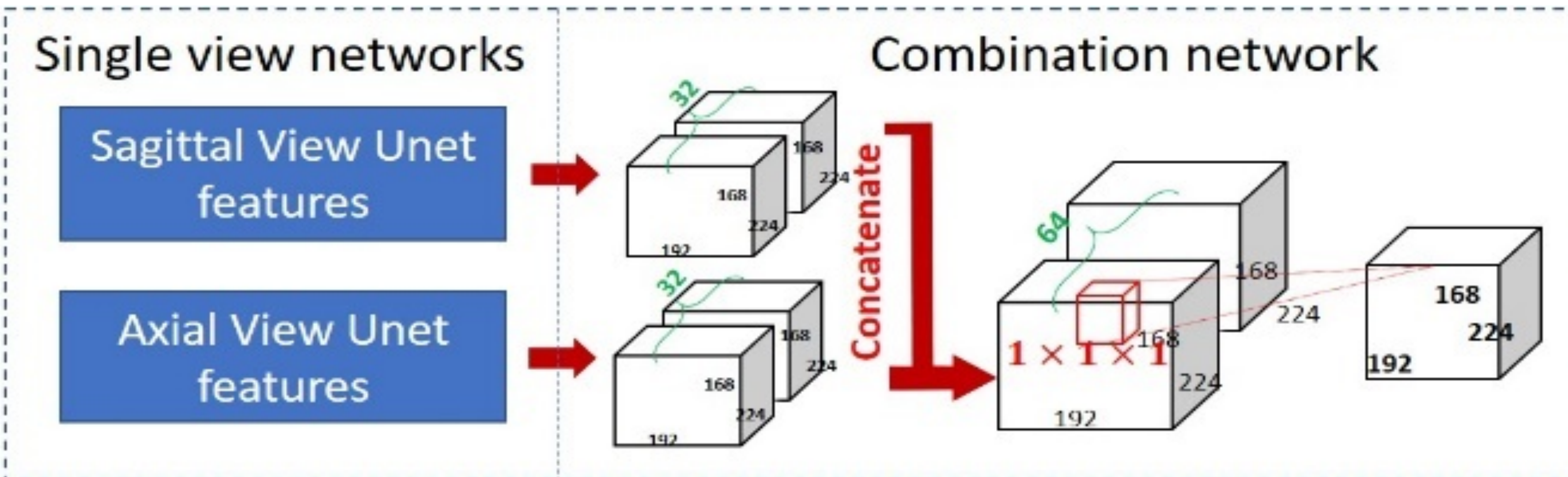


3. Deep Learning Nets 2.5D: 2D at different views → combine

- Axial view: good but with serious stitching blurs
- Sagittal view: overall good
- Coronal view: very bad

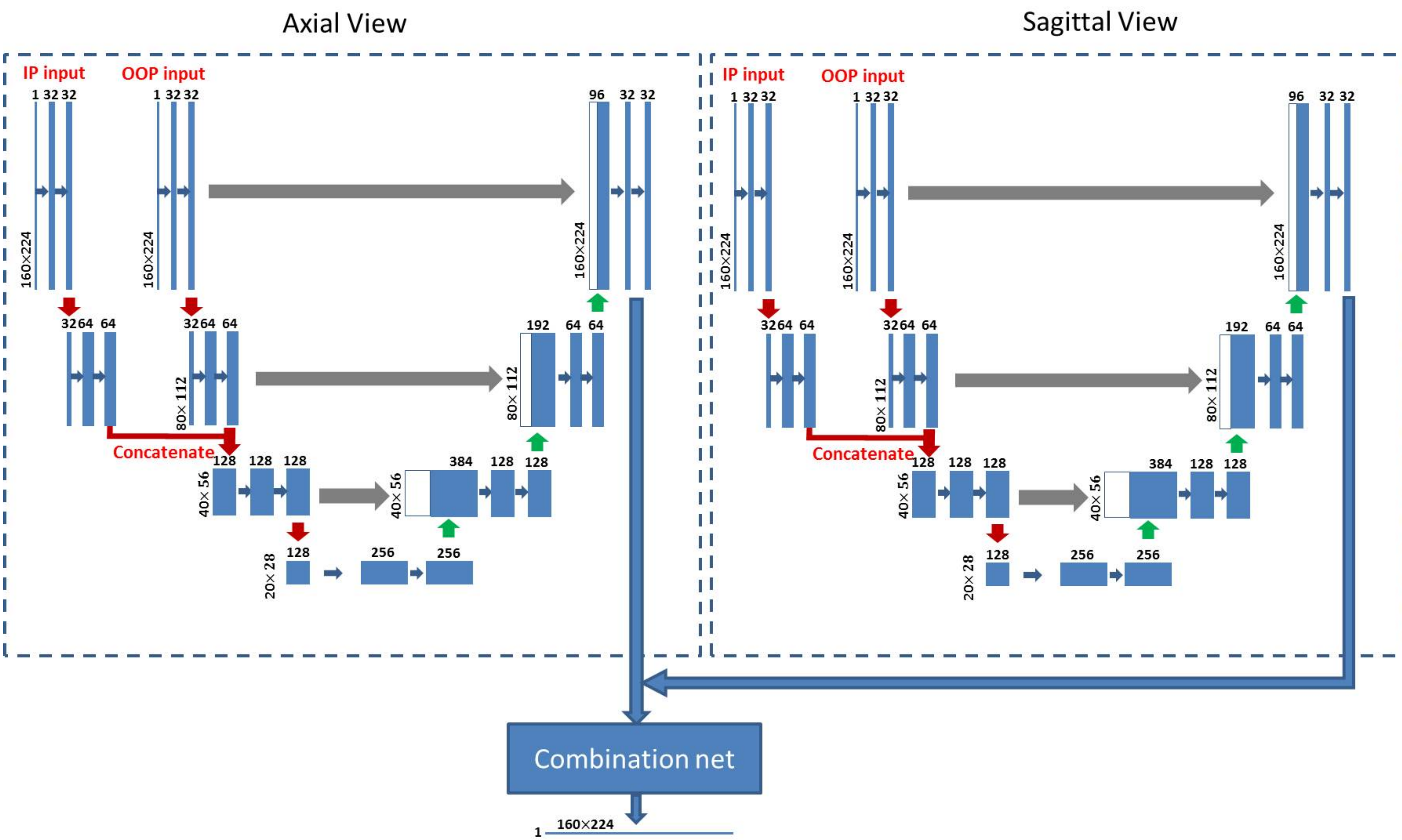
3. Deep Learning Nets 2.5D: 2D at different views ➔ combine

- Combine:



Summary

- Multi-channel Multi-view net with hinge-like smL1 loss




Thank You!