Kid-Net: Convolution Networks for Kidney Vessels Segmentation from CT-Volumes

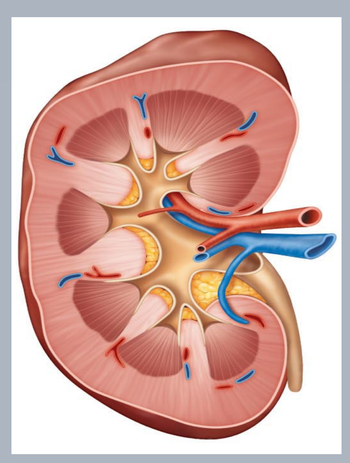
Ahmed Taha, Pechin Lo, Junning Li, Tao Zhao

UNIVERSITY OF MARYLAND

INTUITIVE SURGICAL

Manual medical image annotation is both time-consuming and expensive. Kid-Net reduces kidney vessels segmentation time from matter of hours to minutes. It is trained end-to-end using 3D patches from volumetric CT-images. A complete segmentation for a 512x512x512 CT-volume is obtained within a few minutes (1-2 mins) by stitching the output 3D patches together. Feature down-sampling and up-sampling are utilized to achieve higher classification and localization accuracies. Quantitative and qualitative evaluation results on a challenging testing dataset show Kid-Net competence.

Machine Learning Vessel Segmentation

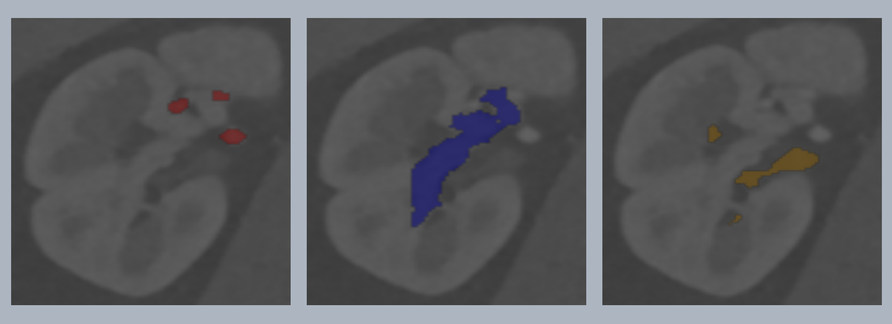


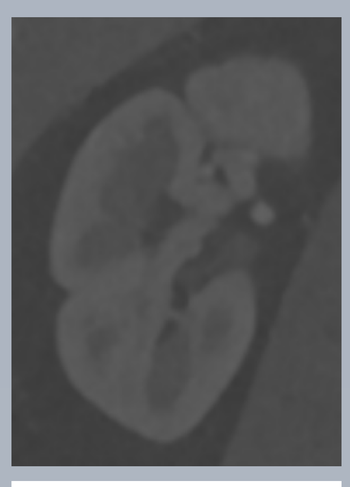
Segment Kidney Anatomy

Artery Vein Ureter (Collecting System)

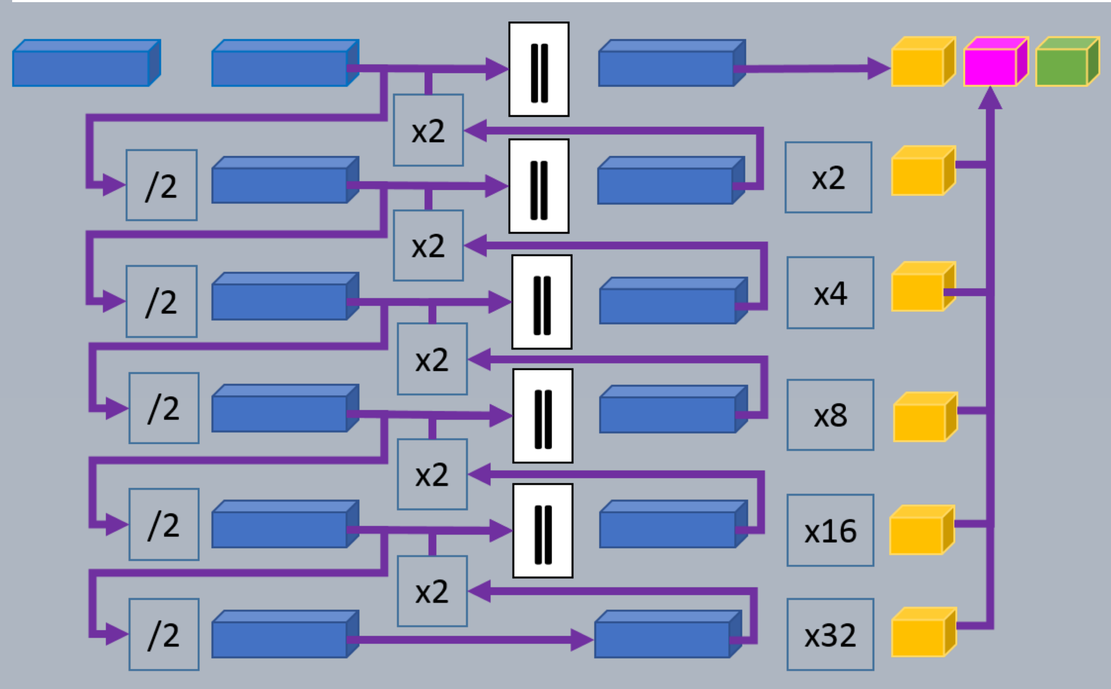
Our main contribution is developing a training schema that handles unbalanced data, reduces false positives and enables high-resolution segmentation with a limited memory budget. These objectives are attained using dynamic weighting, random sampling and 3D patch segmentation.

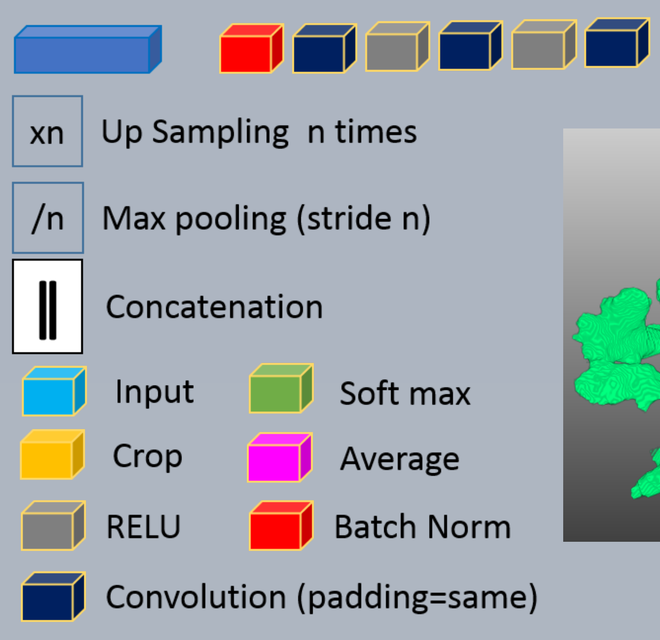
Semantic segmentation is a fundamental step to perform or plan surgical procedures. U-shaped networks managed to train fully 2D convolution network for semantic segmentation in an end-to-end fashion. These architectures have two contradicting phases that carry out complementary tasks. The down-sampling phase detects features, while the up-sampling phase accurately localizes the detected features

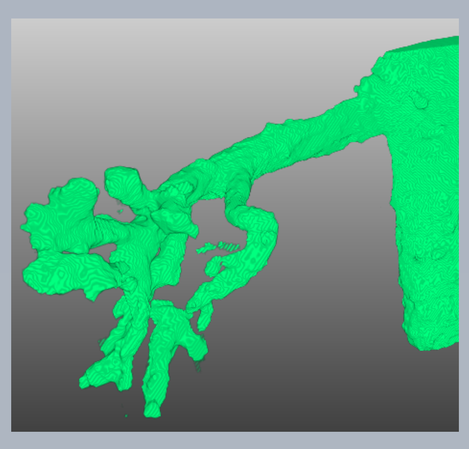




This slice contains artery, vein and collecting system.





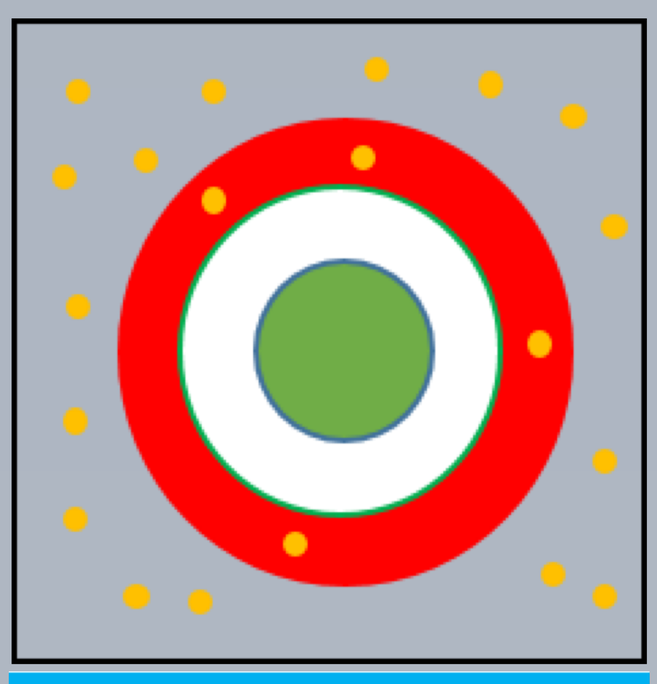


Imbalance data: Two-Fold Approach

1) Dynamic weighting (DW)

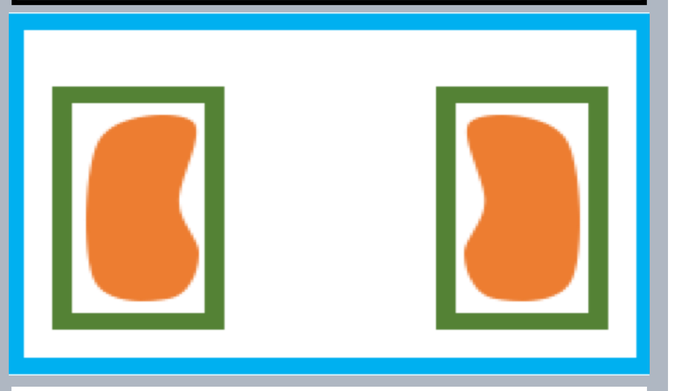
2)Random Background sampling (RS)

Two bands are constructed around kidney vessels using morphological kernel. Misclassifications within the first band (< 2 voxels away from the vessel) are considered marginal errors. In a given patch, the sampled background voxels are equivalent to the foreground vessel volume, where 20% and 80% come from the red band and the volume beyond this band respectively. If a patch is foreground-free, 1% voxels are randomly sampled.



Experiments:

We use volumetric CT-scans from 236 subjects. The average spacing is 0.757x0.757x0.906 mm, with standard deviation 0.075x0.075x0.327 mm. Kid-Net is trained using 3D patches from 99 cases, while 30 and 107 cases are used for validation and testing respectively. Training with foreground-free patches is mandatory. When eliminated, performance degrades because the network learns that every patch has a foreground object, and segments accordingly , which is false. Segmentation results are evaluated using dice-coefficient -- F1 score



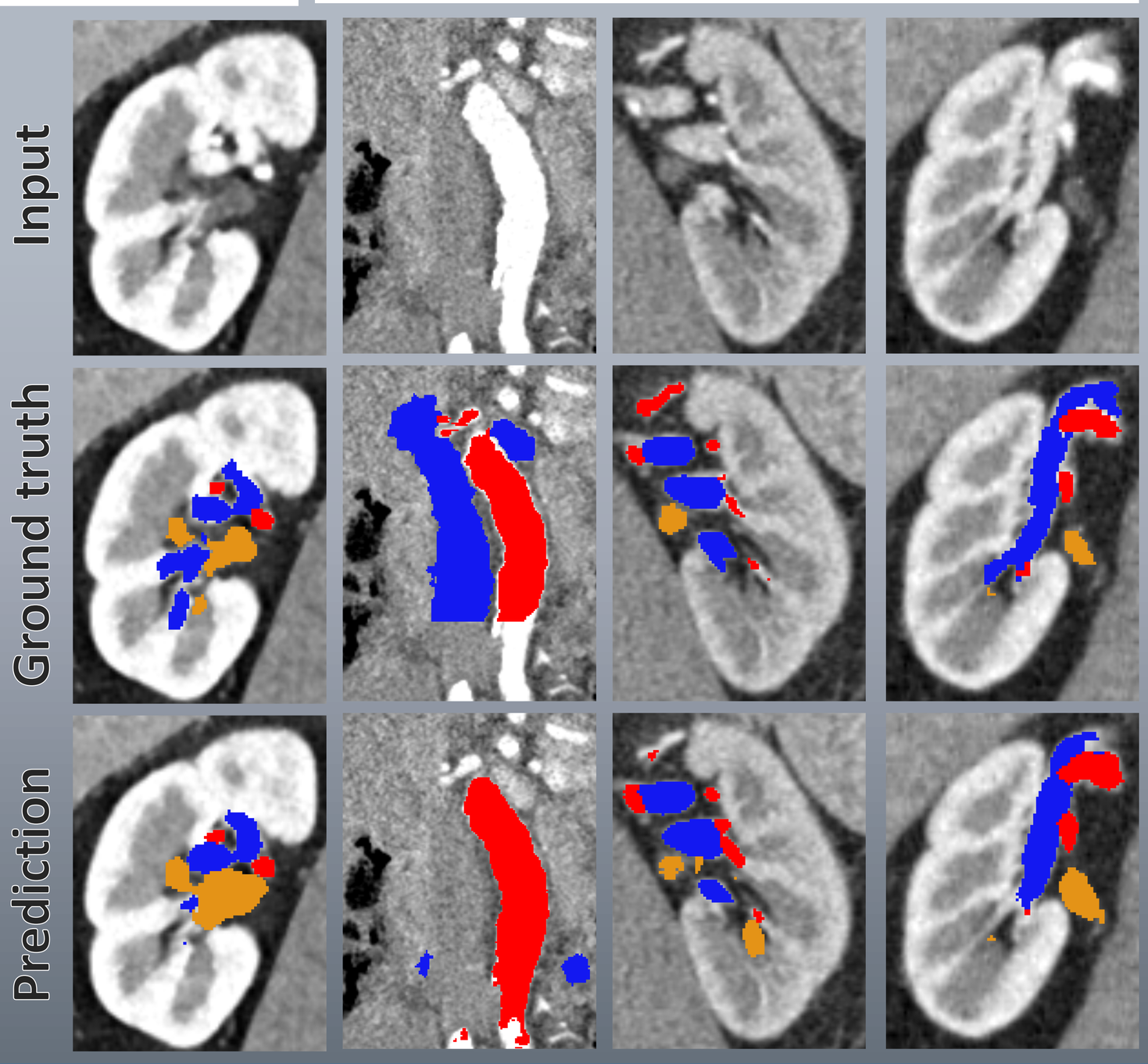
Evaluation regions: The first region is the whole region of interest defined per subject ground truth outlined in blue. The second region is the kidney bounding box outlined in green.

Algorithm: Our proposed moving average procedure assigns voxels dynamic weights (VW_c) based on their classes and patch weight. Patch Weight (PW) is inversely proportional to its foreground volume. Class weight (CW) is inversely proportional to its average volume per patch. Initially $V_c = \frac{1}{n}$ for every class c.

Our settings $\alpha = 0.001, n = 4$.

Require: α : Convergence Rate
Require: P : Current $3D$ patch with axis x, y, z
Require: n : Number of classes (background included)
Require: V_c : Class (c) moving average volume
Require: PW : Current patch weight
Require: CW_c : Class (c) weight
  for all c in classes **do**
    // Measure class (c) volume in patch P
    $V_c(P) == (\sum_x \sum_y \sum_z P(x,y,z) == c)/size(P)$
    // Update class (c) moving average volume
    $V_c = V_c * (1-\alpha) + V_c(P) * \alpha$
    // Update class weight based on its moving average volume
    $CW_c = 1/(n * V_c)$
  end for
  // Set patch weight based on foreground volume
  if P contains background only **then**
    $PW = 1$
  else
    // Foreground volume $\sum_{c=1}^{n-1} V_c(p) < 1$
    $PW = 1 - \log(\sum_{c=1}^{n-1} V_c(p))$
  end if
  $VW_c = PW * CW_c$ (Voxel weight is function of both PW and CW_c)

	Whole ROI		Kidney Bounding Box	
	DW	DW+RS	DW	DW+RS
Artery	0.86	**0.88**	0.72	**0.72**
Vein	**0.59**	0.57	0.60	**0.67**
Ureter	0.32	**0.62**	0.41	**0.63**



Result Analysis

Artery segmentation is the most accurate because all scans are done during arterial phase. This suggests that better vein and ureter segmentations are feasible if venous and waste-out scans are available. The thick aorta boosts artery segmentation F1 score in the whole ROI. In the kidney bounding box, tiny artery vessels become more challenging, and F1 score relatively decreases. The same argument explains vein vessels F1 score. Since arterial scans are used, concealed vena cava penalizes F1 score severely in the whole ROI region.

Among the three kidney vessels, the collecting system is the most challenging. Due to their tiny size, it is difficult to manually annotate or automatically segment. Ureter vessels ground truth annotations are available only within the kidney proximity-- far ureter are less relevant. This explains why ureter F1 score is similar in both evaluation regions. Ureter class is assigned the highest weight due to its relative small size. This leads to a lot of false ureter positives. While random sampling has limited effect on artery and vein F1 score, its merits manifest in ureter segmentation. It boosts segmentation accuracy by 30% and 22% in the whole ROI and kidney bounding box respectively. Thus, It is concluded that both dynamic weights and random sampling are essential to achieve accurate tiny vessels segmentation.

All CT-slices are rendered using soft tissue window--level=40, width=400