Chapter 16
Automatic Pancreas Segmentation Using Coarse-to-Fine Superpixel Labeling

Amal Farag, Le Lu, Holger R. Roth, Jiamin Liu, Evrim Turkbey and Ronald M. Summers

Abstract   Accurate automatic detection and segmentation of abdominal organs from CT images is important for quantitative and qualitative organ tissue analysis, detection of pathologies, surgical assistance as well as computer-aided diagnosis (CAD). In general, the large variability of organ locations, the spatial interaction between organs that appear similar in medical scans and orientation and size variations are among the major challenges of organ segmentation. The pancreas poses these challenges in addition to its flexibility which allows for the shape of the tissue to vastly change. In this chapter, we present a fully automated bottom-up approach for pancreas segmentation in abdominal computed tomography (CT) scans. The method is a four-stage system based on a hierarchical cascade of information propagation by classifying image patches at different resolutions and cascading (segments) superpixels. System components consist of the following: (1) decomposing CT slice images as a set of disjoint boundary-preserving superpixels; (2) computing pancreas class probability maps via dense patch labeling; (3) classifying superpixels by pooling both intensity and probability features to form empirical statistics in cascaded random forest frameworks; and (4) simple connectivity based post-processing. Evaluation of the approach is conducted on a database of 80 manually segmented CT volumes in sixfold cross validation. Our achieved results are comparable, or better to the state-of-the-art methods (evaluated by "leave-one-patient-out"), with a Dice coefficient of 70.7% and Jaccard Index of 57.9%. The computational efficiency of the proposed approach is drastically improved in the order of 6–8 min, compared to other methods of $\geq$10 hours per testing case.

A. Farag · L. Lu (✉) · H.R. Roth · J. Liu · E. Turkbey · R.M. Summers
Department of Radiology and Imaging Sciences, National Institutes of Health
Clinical Center, Bethesda, MD 20837, USA
e-mail: le.lu@nih.gov

R.M. Summers
e-mail: rms@nih.gov

16.1 Introduction

Image segmentation is a key step in image understanding that aims at separating objects within an image into classes, based on object characteristics and a prior information about the surroundings. This also applies to medical image analysis in various imaging modalities. The segmentation of abdominal organs such as the spleen, liver, and pancreas in abdominal computed tomography (CT) scans can be an important input to computer-aided diagnosis (CAD) systems, for quantitative and qualitative analysis and for surgical assistance. In the instance of quantitative imaging analysis of diabetic patients, a requisite critical step for the development of such CAD systems is segmentation specifically of the pancreas. Pancreas segmentation is also a necessary input for subsequent methodologies for pancreatic cancer detection. The literature is rich in methods of automatic segmentation on CT with high accuracies (e.g., Dice coefficients >90%), of other organs such as the kidneys [1], lungs [2], heart [3], and liver [4]. Yet, high accuracy in automatic segmentation of the pancreas remains a challenge. The literature is not as abundant in either single- or multi-organ segmentation setups.

The pancreas is a highly anatomically variable organ in terms of shape and size and the location within the abdominal cavity shifts from patient to patient. The boundary contrast can vary greatly by the amount of visceral fat in the proximity of the pancreas. These factors and others make segmentation of the pancreas very challenging. Figure 16.1 depicts several manually segmented 3D volumes of various patient pancreases to better illustrate the variations and challenges mentioned. From the above observations, we argue that the automated pancreas segmentation problem
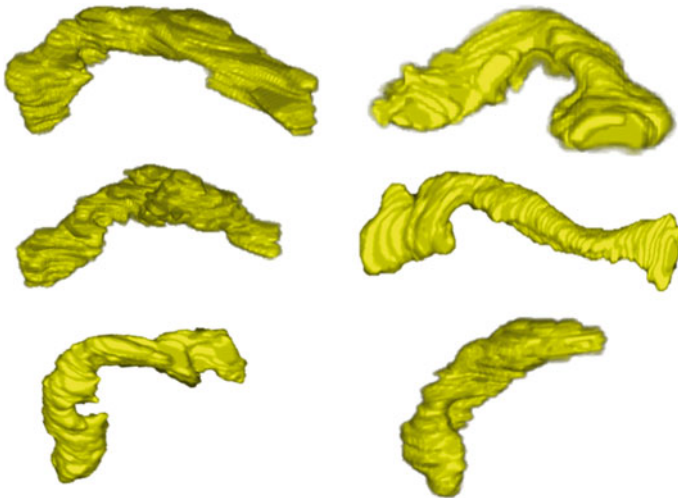

Fig. 16.1 3D manually segmented volumes of six pancreases from six patients. Notice the shape and size variations

should be treated differently, apart from the current organ segmentation literature where statistical shape models are generally used.

In this chapter, a new fully bottom-up approach using image and (deep) patch-level labeling confidences for pancreas segmentation is proposed using 80 single-phase CT patient data volumes. The approach is motivated to improve the segmentation accuracy of *highly deformable organs*, like the pancreas, by leveraging *middle-level representation of image segments*. First, over segmentation of all 2D slices of an input patient abdominal CT scan is obtained as a semi-structured representation known as superpixels. Second, classifying superpixels into two semantic classes of pancreas and non-pancreas is conducted as a multistage feature extraction and random forest (RF) classification process, on the image and (deep) patch-level confidence maps, pooled at the superpixel level. Two cascaded random forest superpixel classification frameworks are presented and compared. Figure 16.2 depicts the overall proposed first framework. Figure 16.9 illustrates the modularized flow charts of both frameworks. Our experimental results are carried out in a sixfold cross-validation manner.
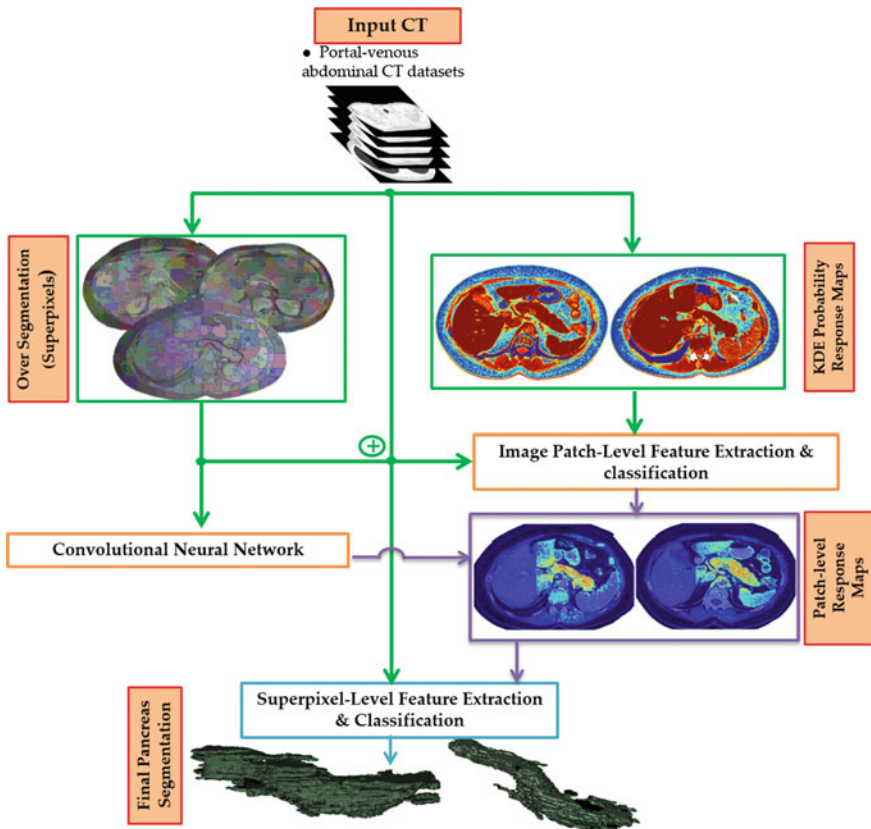

Fig. 16.2   Overall pancreas segmentation framework via dense image patch labeling

Our system runs at about two orders of magnitude more computationally efficiently to process a new testing case than the atlas registration based approaches [5–10]. The obtained results are comparable, or better than the state-of-the-art methods (evaluated by "leave-one-patient-out"), with a Dice coefficient of 70.7% and Jaccard Index of 57.9%. Under the same sixfold cross validation, our bottom-up segmentation method significantly outperforms its "multi-atlas registration and joint label fusion" (MALF) counterpart (based on our implementation using [11, 12]): Dice coefficients $70.7 \pm 13.0\%$ versus $52.51 \pm 20.84\%$. Additionally, another bottom-up supervoxel based multi-organ segmentation without registration in 3D abdominal CT images is also investigated [13] in a similar spirit, for demonstrating this methodological synergy.

16.2   Previous Literature

The organ segmentation literature can be divided into two broad categories: top-down and bottom-up approaches. In top-down approaches, a priori knowledge such as atlas(es) and/or shape models of the organ are generated and incorporated into the framework via learning based shape model fitting [3, 4] or volumetric image registration [7, 8, 10]. For bottom-up approaches segmentation is performed by local image similarity grouping and growing or pixel, superpixel/supervoxel-based labeling [14, 15] since direct representations of the organ is not incorporated. Generally speaking, top-down methods are targeted for organs which can be modeled well by statistical shape models [3] whereas bottom-up representations are more effective for highly non-Gaussian shaped [14, 15] or pathological organs.

Previous literature on pancreas segmentation from CT images have been dominated by top-down approaches which rely on atlas-based approaches or statistical shape modeling or both [5–10].

- Shimizu et al. [5] utilize three-phase contrast enhanced CT data which are first registered together for a particular patient and then registered to a reference patient by landmark-based deformable registration. The spatial support area of the abdominal cavity is reduced by segmenting the liver, spleen, and three main vessels associated with location interpretation of the pancreas (i.e., splenic, portal, and superior mesenteric veins). Coarse-to-fine pancreas segmentation is performed by using generated patient-specific probabilistic atlas guided segmentation followed by intensity-based classification and post-processing. Validation of the approach was conducted on 20 multi-phase datasets resulting in a Jaccard of 57.9%.
- Okada et al. [6] perform multi-organ segmentation by combining inter-organ spatial interrelations with probabilistic atlases. The approach incorporated various a priori knowledge into the model that includes shape representations of seven organs. Experimental validation was conducted on 28 abdominal contrast-enhanced CT datasets obtaining an overall volume overlap of Dice index 46.6% for the pancreas.

- Chu et al. [8] present an automated multi-organ segmentation method based on spatially divided probabilistic atlases. The algorithm consists of image-space division and a multi-scale weighting scheme to deal with the large differences among patients in organ shape and position in local areas. Their experimental results show that the liver, spleen, pancreas, and kidneys can be segmented with Dice similarity indices of 95.1, 91.4, 69.1, and 90.1%, respectively, using 100 annotated abdominal CT volumes.
- Wolz et al. [7] may be considered the state-of-the-art result thus far for single-phase pancreas segmentation. The approach is a multi-organ segmentation approach that combines hierarchical weighted subject-specific atlas-based registration and patch-based segmentation. Post-processing is in the form of optimized graph-cuts with a learned intensity model. Their results in terms of a Dice overlap for the pancreas is 69.6% on 150 patients and 58.2% on a subpopulation of 50 patients.
- Recent work by Wang et al. [10] proposes a patch-based label propagation approach that uses relative geodesic distances. The approach can be considered a start to developing some bottom-up component for segmentation, where affine registration between dataset and atlases were conducted followed by refinement using the patch-based segmentation to reduce misregistrations and instances of high anatomy variability. The approach was evaluated on 100 abdominal CT scans with an overall Dice of 65.5% for the pancreas segmentation.

The default experimental setting in many of the atlas-based approaches [5–10] is conducted in a "leave-one-patient-out" or "leave-one-out" (LOO) criterion for up to N = 150 patients. In the clinical setting, leave-one-out based dense volume registration (from all other N-1 patients as atlas templates) and label fusion process may be computationally impractical (10+ hours per testing case). More importantly, it does not scale up easily when large-scale datasets are present. On the other hand, efficient cascade classifiers have been studied in both computer vision and medical image analysis problems [16–18], with promising results.

16.3 Methods

In this section, the components of our overall algorithm flow (shown in Fig. 16.2) are first addressed (Sects. 16.3.1 and 16.3.2). The method extensions on exploiting sliding-window CNN-based dense image patch labeling and framework variations are described in Sects. 16.3.3 and 16.3.4.

16.3.1 Boundary-Preserving Over-segmentation

Over-segmentation occurs when images (or more generally grid graphs) are segmented or decomposed into smaller perceptually meaningful regions, "superpixels".

Within a superpixel, pixels carry similarities in color, texture, intensity, etc., and generally align with image edges rather than rectangular patches (i.e., superpixels can be irregular in shape and size). In the computer vision literature, numerous approaches have been proposed for superpixel segmentation [19–23]. Each approach has its drawbacks and advantages but three main properties are generally examined when deciding the appropriate method for an application as discussed in [20]: (1) adherence to image boundaries; (2) computationally fast, ease of usage and memory efficient; especially when computational complexity reduction is of importance and (3) improvement on both quality and speed of the final segmentation.

Superpixel methods fall under two main broad categories: graph-based (e.g., SLIC [19], entropy rate [21] and [22]) and gradient ascent methods (e.g., watershed [23] and mean shift [24]). In terms of computational complexity, [22, 23] are relatively fast in $O(MlogM)$ complexity where M is the number of pixels or voxels in the image or grid graph. Mean shift [24] and normalized cut [25] are $O(M^2)$, or $O(M^{\frac{3}{2}})$, respectively. Simple linear iterative clustering (SLIC) [19] is both fast and memory efficient. In our work, evaluation and comparison among three graph-based superpixel algorithms (i.e., SLIC [19, 20], efficient graph-based [22] and Entropy rate [21]) and one gradient ascent method (i.e., watershed [23]) are conducted, considering the three criterion in [20]. Figure 16.3 shows sample superpixel results using the SLIC approach. The original CT slices and cropped zoomed-in pancreas superpixel regions are demonstrated. The boundary recall, a typical measurement used in the literature, to indicate how many "true" edge pixels of the ground-truth object segmentation are within a pixel range from the superpixels (i.e., object-level edges are recalled by superpixel boundaries). High boundary recall indicates minimal true edges were neglected. Figure 16.4 shows sample quantitative results. Based on Fig. 16.3, high boundary recalls, within the distance ranges between 1 and 6 pixels from the semantic pancreas ground-truth boundary annotation are obtained using the SLIC approach. The watershed approach provided the least promising results for usage in the pancreas, due to the lack of conditions in the approach, to utilize boundary information in conjunction with intensity information as implemented in graph-based approaches. The superpixel number range per axial image is constrained $\in [100, 200]$ to make a good trade-off on superpixel dimensions or sizes.

The overlapping ratio r of the superpixel versus the ground-truth pancreas annotation mask is defined as the percentage of pixels/voxels inside each superpixel that are annotated as pancreas. By thresholding on r, say if $r > \tau$ the superpixel will be labeled as pancreas and otherwise as background, we can obtain the pancreas segmentation results. When $\tau = 0.50$, the achieved mean Dice coefficient is $81.2 \pm 3.3\%$ which is referred as the "Oracle" segmentation accuracy since computing r would require to know the ground-truth segmentation. This is also the upper bound segmentation accuracy for our superpixel labeling or classification framework. $81.2 \pm 3.3\%$ is significantly higher and numerically more stable (in standard deviation) than previous state-of-the-art methods [5, 7–10], to provide considerable improvement space of our work. Note that both the choices of SLIC and $\tau = 0.50$ are calibrated using a subset of 20 scans. We find there is no need to evaluate different superpixel generation methods/parameters and τs as "model selection" using the
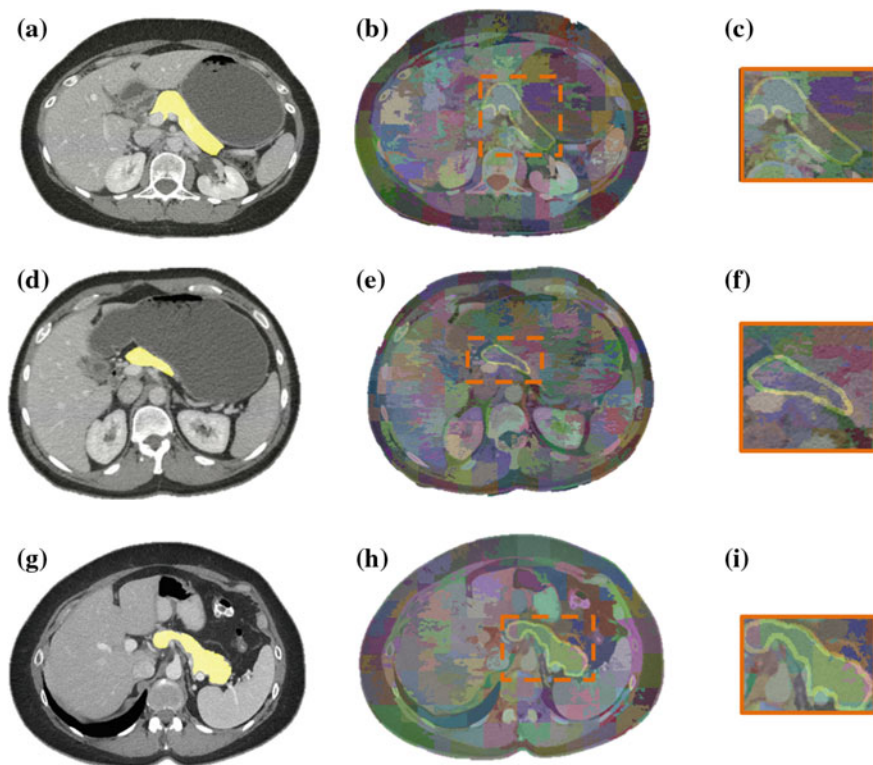
Fig. 16.3 Sample superpixel generation results from the SLIC method [19]. *First column* depicts different slices from different patient scans with the ground-truth pancreas segmentation in *yellow* (**a**, **d** and **g**). The *second column* depicts the over segmentation results with the pancreas contours superimposed on the image (**b**, **e** and **h**). *Last*, (**c**) (**f**) and (**i**) show zoomed-in areas of the pancreas superpixel results from **b**, **e** and **h**

Fig. 16.4 Superpixels boundary recall results evaluated on 20 patient scans (Distance in millimeters). The watershed method [23] is shown in *red*, efficient graph [22] in *blue* while the SLIC [19] and the Entropy rate [21] based methods are depicted in *cyan* and *green*, respectively. The *red line* represents the 90% marker
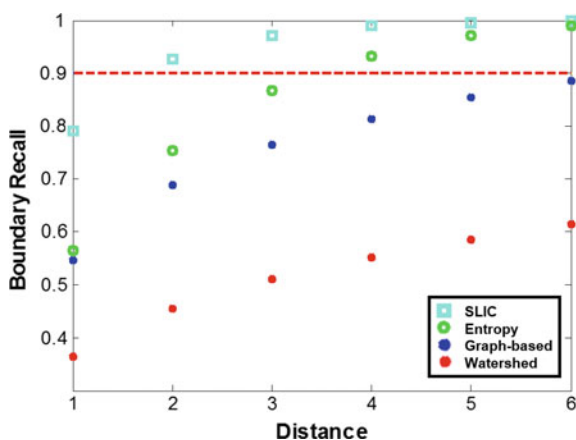
training folds in each round of sixfold cross validation. This superpixel calibration procedure is generalized well to all our datasets. Voxel-level pancreas segmentation can be propagated from superpixel-level classification and further improved by efficient narrow-band level-set based curve evolution [26], or the learned intensity model based graph-cut [7].

16.3.2 Patch-Level Visual Feature Extraction and Classification: P^{RF}

Feature extraction is a form of object representation that aims at capturing the important shape, texture, and other salient features that allow distinctions between the desired object (i.e., pancreas) and the surrounding to be made. In this work a total of 46 patch-level image features to depict the pancreas and its surroundings are implemented. The overall 3D abdominal body region per patient is first segmented and identified using a standard table-removal procedure where all voxels outside the body are removed.

(1) To describe the texture information, we adopt the Dense Scale-Invariant Feature transform (dSIFT) approach [27] which is derived from the SIFT descriptor [28] with several technical extensions. The publicly available VLFeat implementation of the dSIFT is employed [27]. Figure 16.5 depicts the process implemented on a sample image slice. The descriptors are densely and uniformly extracted from
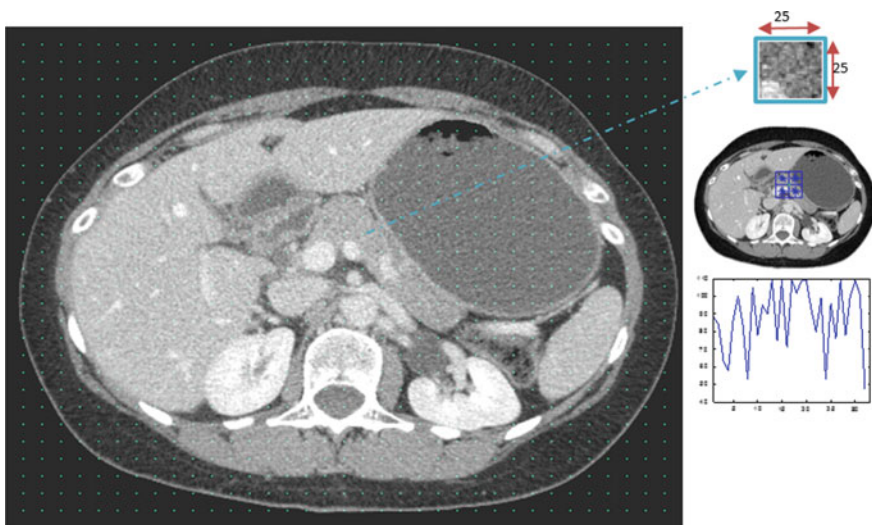

Fig. 16.5 Sample slice with center positions superimposed as *green dots*. The 25 × 25 image patch and corresponding D-SIFT descriptors are shown to the *right* of the original image

image grids with inter-distances of 3 pixels. The patch center position are shown as the green points superimposed on the original image slice. Once the positions are known, the dSIFT is computed with the geometry of [2 × 2] bins and bin size of 6 pixels, which results in a 32 dimensional texture descriptor for each image patch. The image patch size in this work is fixed at 25 × 25 which is a trade-off between computational efficiency and description power. Empirical evaluation of the image patch size is conducted for the size range of 15–35 pixels using a small subsampled dataset for classification, as described later. Stable performance statistics are observed and quantitative experimental results using the default patch size of 25 × 25 pixels are reported.

(2) A second feature group using the voxel intensity histograms of the ground-truth pancreas and the surrounding CT scans is built in the class-conditional probability density function (PDF) space. A kernel density estimator (KDE[1]) is created using the voxel intensities from a subset of randomly selected patient CT scans. The KDE represents the CT intensity distributions of the positive $\{X^+\}$ and negative class $\{X^-\}$ of pancreas and non-pancreas voxels CT image information. All voxels containing pancreas information are considered in the positive sample set, yet, since negative voxels far outnumber the positive only 5% of the total number from each CT scan (by random resampling) is considered. Let, $\{X^+\} = \left(h_1^+, h_2^+, \ldots, h_n^+\right)$ and $\{X^-\} = \left(h_1^-, h_2^-, \ldots, h_m^-\right)$ where h_n^+ and h_m^- represent the intensity values for the positive and negative pixel samples for all 26 patient CT scans over the entire abdominal CT Hounsfield range. The kernel density estimators $f^+(X^+) = \frac{1}{n}\sum_{i=1}^n K\left(X^+ - X_i^+\right)$ and $f^-(X^-) = \frac{1}{m}\sum_{j=1}^m K\left(X^- - X_j^-\right)$ are computed where $K()$ is assumed to be a Gaussian kernel with optimal computed bandwidth, for this data, of 3.039. Kernel sizes or bandwidth may be selected automatically using 1D Likelihood-based search, as provided by the used KDE toolkit. The normalized likelihood ratio is calculated which becomes a probability value as a function of intensity in the range of $H = [0 : 1 : 4095]$. Thus, the probability of being considered pancreas is formulated as $y^+ = \frac{(f^+(X^+))}{(f^+(X^+)+f^-(X^-))}$. This function is converted as a precomputed lookup table over $H = [0 : 1 : 4095]$, which allows very efficient $O(1)$ access time.

(3) Utilizing first the KDE probability response maps above and the superpixel CT masks described in Sect. 16.3.1, as underlying supporting masks to each image patch, the same KDE response statistics within the intersected subregions, P' of P, are extracted. The idea is that an image patch, P, may be divided into more than one superpixel. This set of statistics is calculated with respect to the most representative superpixel (that covers the patch center pixel). In this manner, object boundary-preserving intensity features are obtained.

(4) The final two features for each axial slice (in the patient volumes) are the normalized relative x-axis and y-axis positions $\epsilon[0, 1]$, computed at each image patch
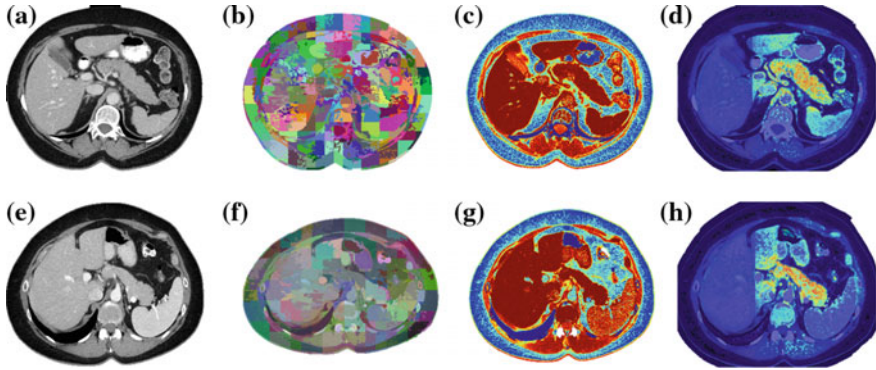

Fig. 16.6 Two sample slices from different patients are shown in **a** and **e**. The corresponding super-pixels segmentation (**b**, **f**), KDE probability response maps (**c**, **g**) and RF patch-level probability response maps (**d**, **h**) are shown. In **c**, **g** and **d**, **h**, *red* represents highest probabilities. In **d**, **h** the *purple color* represents areas where probabilities are so small and can be deemed insignificant areas of interest

center against the segmented body region (self-normalized[2] to patients with different body masses to some extent). Once all of the features are concatenated, a total of 46 image patch-level features per superpixel are used to train a random forest (RF) classifier C_p. Image patch labels are obtained by directly borrowing the class information of their patch center pixels, based on the manual segmentation.

Sixfold cross validation for RF training is carried out. Response maps are computed for the image patch-level classification and dense labeling. Figure 16.6d, h show sample illustrative slices from different patients. High probability corresponding to the pancreas is represented by the red color regions (the background is blue). The response maps (denoted as P^{RF}) allow several observations to be made. The most interesting is that the relative x and y positions as features allow for clearer spatial separation of positive and negative regions, via internal RF feature thresholding tests on them. The trained RF classifier is able to recognize the negative class patches residing in the background, such as liver, vertebrae and muscle using spatial location cues. In Fig. 16.6d, h implicit vertical and horizontal decision boundary lines can be seen in comparison to Fig. 16.6c, g. This demonstrates the superior descriptive and discriminative power of the feature descriptor on image patches (P and P') than single pixel intensities. Organs with similar CT values are significantly depressed in the patch-level response maps.

In summary, SIFT and its variations, e.g., D-SIFT have shown to be informative, especially through spatial pooling or packing [29]. A wide range of pixel-level correlations and visual information per image patch is also captured by the rest of 14

---

[2]The axial reconstruction CT scans in our study have largely varying ranges or extends in the z-axis. If some anatomical landmarks, such as the bottom plane of liver, the center of kidneys, can be provided automatically, the anatomically normalized z-coordinate positions for superpixels can be computed and used as an additional spatial feature for RF classification.
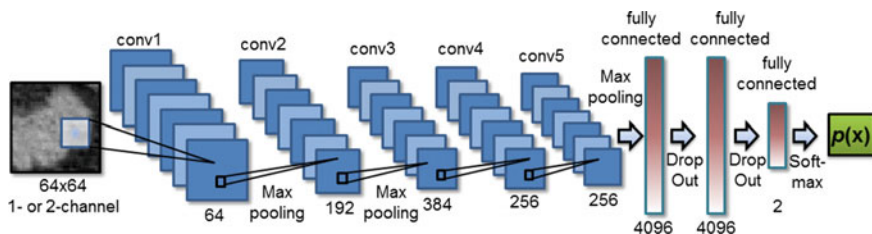
Fig. 16.7 The proposed CNN model architecture is composed of five convolutional layers with max pooling and two fully connected layers with DropOut [30] connections. A final two-way softmax layer gives a probability $p(x)$ of "pancreas" and "non-pancreas" per data sample (or image patch). The number and model parameters of convolutional filters and neural network connections for each layer are as shown

defined features. Both good classification specificity and recall have been obtained in cross validation using Random Forest implementation of 50 trees and the minimum leaf size set as 150 (i.e., using the *treebagger*($\bullet$) function in Matlab).

16.3.3   Patch-Level Labeling via Deep Convolutional Neural Network: P^{CNN}

In this work, we use Convolutional Neural Network (CNN, or ConvNet) with a standard architecture for binary image patch classification. Five layers of convolutional filters first compute, aggregate, and assemble the low level image features to more complex ones, in a layer-by-layer fashion. Other CNN layers perform max-pooling operations or consist of fully connected neural network layers. The CNN model we adopted ends with a final two-way softmax classification layer for "pancreas" and "non-pancreas" classes (refer to Fig. 16.7). The fully connected layers are constrained using "DropOut" in order to avoid over-fitting in training where each neuron or node has a probability of 0.5 to be reset with a 0-valued activation. DropOut is a method that behaves as a co-adaption regularizer when training the CNN [30]. In testing, no DropOut operation is needed. Modern GPU acceleration allows efficient training and run-time testing of the deep CNN models. We use the publicly available code base of *cuda-convnet2*.[3]

To extract dense image patch response maps, we use a straight-forward sliding-window approach that extracts 2.5D image patches composed of axial, coronal, and sagittal planes at any image positions (see Fig. 16.8). Deep CNN architecture can encode large-scale image patches (even the whole 224×224 pixel images [31, 32]) very efficiently and no hard crafted image features are required any more. In this work, the dimension of image patches for training CNN is 64×64 pixels which is

---

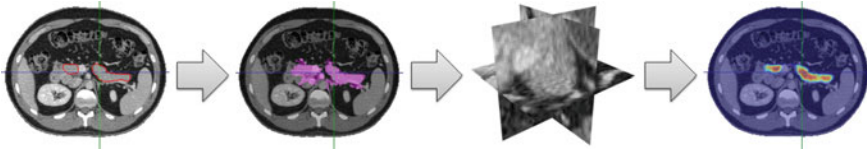[3]https://code.google.com/p/cuda-convnet2.

Fig. 16.8 Axial CT slice of a manual (gold standard) segmentation of the pancreas. From *Left* to *Right*, there are the ground-truth segmentation contours (in *red*); RF-based coarse segmentation $\{S_{RF}\}$; a 2.5D input image patch to CNN and the deep patch labeling result using CNN

significantly larger than 25×25 in Sect. 16.3.2. The larger spatial scale or context is generally expected to achieve more accurate patch labeling quality. For efficiency reasons, we extract patches every ℓ voxels for CNN feedforward evaluation and then apply nearest neighbor interpolation to estimate the values at skipped voxels. In our empirical testing, simple nearest neighbor interpolation seems sufficient due to the high quality of deep CNN probability predictions. Three examples of dense CNN based image patch labeling are demonstrated in Fig. 16.10. We denote the CNN model generated probability maps as P^{CNN}.

The computational expense of deep CNN patch labeling per patch (in a sliding-window manner) is still higher than Sect. 16.3.2. In practice, dense patch labeling by P^{RF} runs exhaustively at 3 pixel interval but P^{CNN} are only evaluated at pixel locations that pass the first stage of a cascaded random forest superpixel classification framework. This process is detailed in Sect. 16.3.4 where C_{SP}^1 is operated at a high recall (close to 100%) and low specificity mode to minimize the false negative rate (FNR) as the initial layer of cascade. The other important reason for doing so is to largely alleviate the training unbalance issue for P^{CNN} in C_{SP}^3. After this initial pruning, the number ratio of non-pancreas versus pancreas superpixels changes from >100 to ~ 5. The similar treatment is employed in our recent work [33] where all "Regional CNN" (R-CNN) based algorithmic variations [34] for pancreas segmentation is performed after a superpixel cascading.

16.3.4 Superpixel-Level Feature Extraction, Cascaded Classification, and Pancreas Segmentation

In this section, we trained three different superpixel-level random forest classifiers of C_{SP} 1, C_{SP} 2, and C_{SP} 3. These three classifier components further formed two cascaded RF classification frameworks (F-1, F-2), as shown in Fig. 16.9. The superpixel labels are inferred from the overlapping ratio r (defined in Sect. 16.3.1) between the superpixel label map and the ground-truth pancreas mask. If $r \geq 0.5$, the superpixel is positive while if $r \leq 0.2$, the superpixel is assigned as negative. For the rest of superpixels that fall within $0.2 < r < 0.5$ (a relatively very small portion/subset of all superpixels), they are considered ambiguous and not assigned a label and as such not used in training.
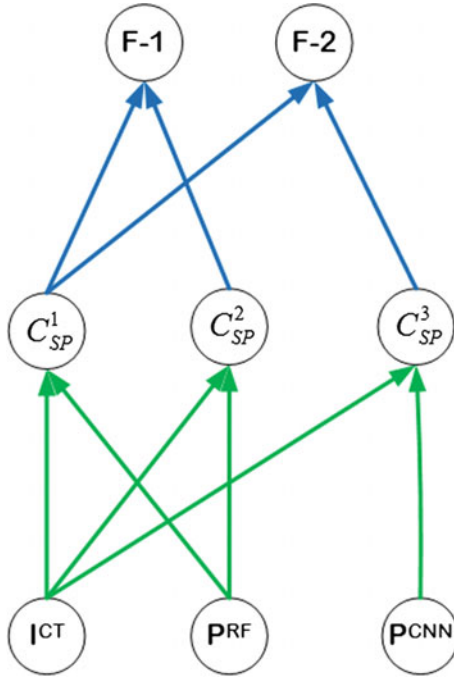
Fig. 16.9 The flow chart of input channels and component classifiers to form the overall frameworks 1 (F-1) and 2 (F-2). I^{CT} indicates the original CT image channel; P^{RF} represents the probability response map by RF-based patch labeling in Sect. 16.3.2 and P^{CNN} from deep CNN patch classification in Sect. 16.3.3, respectively. Superpixel level random forest classifier C_{SP}^1 is trained with all positive and negative superpixels in I^{CT} and P^{RF} channels; C_{SP}^2 and C_{SP}^3 are learned using only "hard negatives" and all positives, in the $I^{CT} \bigcup P^{RF}$ or $I^{CT} \bigcup P^{CNN}$ channels, respectively. Forming $C_{SP}^1 \mapsto C_{SP}^2$, or $C_{SP}^1 \mapsto C_{SP}^3$ into two overall cascaded models results in frameworks F-1 and F-2

Training C_{SP}^1 utilizes both the original CT image slices (I^{CT} in Fig. 16.9) and the probability response maps (P^{RF}) via the handcrafted feature based patch-level classification (i.e., Sect. 16.3.2). The 2D superpixel supporting maps (i.e., Sect. 16.3.1) are used for feature pooling and extraction on a superpixel level. The CT pixel intensity/attenuation numbers and the per-pixel pancreas class probability response values (from dense patch labeling of P^{PF} or P^{CNN} later) within each superpixel are treated as two empirical unordered distributions. Thus our superpixel classification problem is converted as modeling the difference between empirical distributions of positive and negative classes. We compute (1) simple statistical features of the first–fourth order statistics such as mean, std, skewness, kurtosis [35] and (2) histogram-type features of eight percentiles (20, 30, ..., 90%), per distribution in intensity or P^{RF} channel, respectively. Once concatenated, the resulted 24 features for each superpixel instance is fed to train random forest classifiers.

Due to the highly unbalanced quantities between foreground (pancreas) super-pixels and background (the rest of CT volume) superpixels, a two-tiered cascade of random forests is exploited to address this type of rare event detection problem [36]. In a cascaded classification, C_{SP}^1 once trained is applied exhaustively on scanning all superpixels in an input CT volume. Based on the receiver operating characteristic (ROC) curves for C_{SP}^1, we can safely reject or prune 97% negative superpixels while maintaining nearly ∼100% recall or sensitivity. The remained 3% negatives, often referred as "hard negatives" [36], along with all positives are employed to train the second C_{SP}^2 in the same feature space. Combining C_{SP}^1 and C_{SP}^2 is referred to as Framework 1 (F-1) in the subsequent sections.

Similarly, we can train a random forest classifier C_{SP}^3 by replacing C_{SP}^2's feature extraction dependency on the P^{RF} probability response maps, with the deep CNN patch classification maps of P^{CNN}. The same 24 statistical moments and percentile features per superpixel, from two information channels I^{CT} and P^{CNN}, are extracted to train C_{SP}^3. Note that the CNN model that produces P^{CNN} is trained with the image patches sampled from only "hard negative" and positive superpixels (aligned with the second-tier RF classifiers C_{SP}^2 and C_{SP}^3). For simplicity, P^{RF} is only trained once with all positive and negative image patches. This will be referred to as Framework 2 (F-2) in the subsequent sections. F-1 only use P^{RF} whereas F-2 depends on both P^{RF} and P^{CNN} (with a little extra computational cost).

The flow chart of frameworks 1 (F-1) and 2 (F-2) is illustrated in Fig. 16.9. The two-level cascaded random forest classification hierarchy is found empirically to be sufficient (although a deeper cascade is possible) and implemented to obtain F-1: C_{SP}^1 and C_{SP}^2, or F-2: C_{SP}^1 and C_{SP}^3. The binary 3D pancreas volumetric mask is obtained by stacking the binary superpixel labeling outcomes (after C_{SP}^2 in F-1 or C_{SP}^3 in F-2) for each 2D axial slice, followed by 3D connected component analysis implemented in the end. By assuming the overall pancreas connectivity of its 3D shape, the largest 3D connected component is kept as the final segmentation. The binarization thresholds of random forest classifiers in C_{SP}^2 and C_{SP}^3 are calibrated using data in the training folds in sixfold cross validation, via a simple grid search. In [33], standalone *Patch-ConvNet* dense probability maps (without any post-processing) are processed for pancreas segmentation after using (F-1) as an initial cascade. The corresponding pancreas segmentation performance is not as accuracy as (F-1) or (F-2).

16.4   Data and Experimental Results

16.4.1   Imaging Data

80 3D abdominal portal-venous contrast-enhanced CT scans (∼70 s after intravenous contrast injection) acquired from 53 male and 27 female subjects are used in our study for evaluation. Seventeen of the subjects are from a kidney donor transplant list of healthy patients that have abdominal CT scans prior to nephrectomy. The remaining

63 patients are randomly selected by a radiologist from the Picture Archiving and Communications System (PACS) on the population that has neither major abdominal pathologies nor pancreatic cancer lesions. The CT datasets are obtained from National Institutes of Health Clinical Center. Subjects range in the age from 18 to 76 years with a mean age of 46.8 ± 16.7. Scan resolution has 512×512 pixels (varying pixel sizes) with slice thickness ranging from 1.5 to 2.5 mm on Philips and Siemens MDCT scanners. The tube voltage is 120 kVp. Manual ground-truth segmentation masks of the pancreas for all 80 cases are provided by a medical student and verified/modified by a radiologist.

16.4.2 Experiments

Experimental results are assessed using sixfold cross validation, as described in Sects. 16.3.2 and 16.3.4. Several metrics to evaluate the accuracy and robustness of the methods are computed. The Dice similarity index which interprets the overlap between two sample sets, $SI = 2(|A \cap B|)/(|A| + |B|)$ where A and B refer to the algorithm output and manual ground-truth 3D pancreas segmentation, respectively. The Jaccard index (JI) is another statistic used to compute similarities between the segmentation result against the reference standard, $JI = (|A \cap B|)/(|A \cup B|)$, called "intersection over union" in the PASCAL VOC challenges [37, 38]. The volumetric recall (i.e. sensitivity) and precision values are also reported (Fig. 16.10).

Next, the pancreas segmentation performance evaluation is conducted in respect to the total number of patient scans used for the training and testing phases. Using our framework F1 on 40, 60 and 80 (i.e., 50, 75, and 100% of the total 80 datasets) patient scans, the Dice, JI, Precision, and Recall are computed under sixfold cross validation. Table 16.1 shows the computed results using image patch-level features and multi-level classification (i.e., performing C_{SP}^1 and C_{SP}^2 on I^{CT} and P^{RF}) and how performance changes with the additions of more patients data. Steady improvements of ~4% in the Dice coefficient and ~5% for the Jaccard index are observed, from 40 to 60, and 60–80. Figure 16.11 illustrates some sample final pancreas segmentation results from the 80 patient execution for two different patients. The results are divided into three categories: good, fair, and poor. The good category refers to the computed Dice coefficient above 90% (of 15 patients), fair result as $50\% \leq Dice \geq 90\%$ (49 patients) and poor for Dice $<50\%$ (16 patients).

Then, we evaluate the difference of the proposed F-1 versus F-2 on 80 patients, using the same four metrics (i.e., Dice, JI, precision, and recall). Table 16.1 shows the comparison results. The same sixfold cross-validation criterion is employed so that direct comparisons can be made. From the table, it can be seen that about 2% increase in the Dice coefficient was obtained by using F-2, but the main improvement can be noticed in the minimum values (i.e., the lower performance bound) for each of the metrics. Usage of deep patch labeling prevents the case of no pancreas segmentation while keeping slightly higher mean precision and recall values. The standard deviations also dropped nearly 50% comparing F-1 to F-2 (from 25.6 to
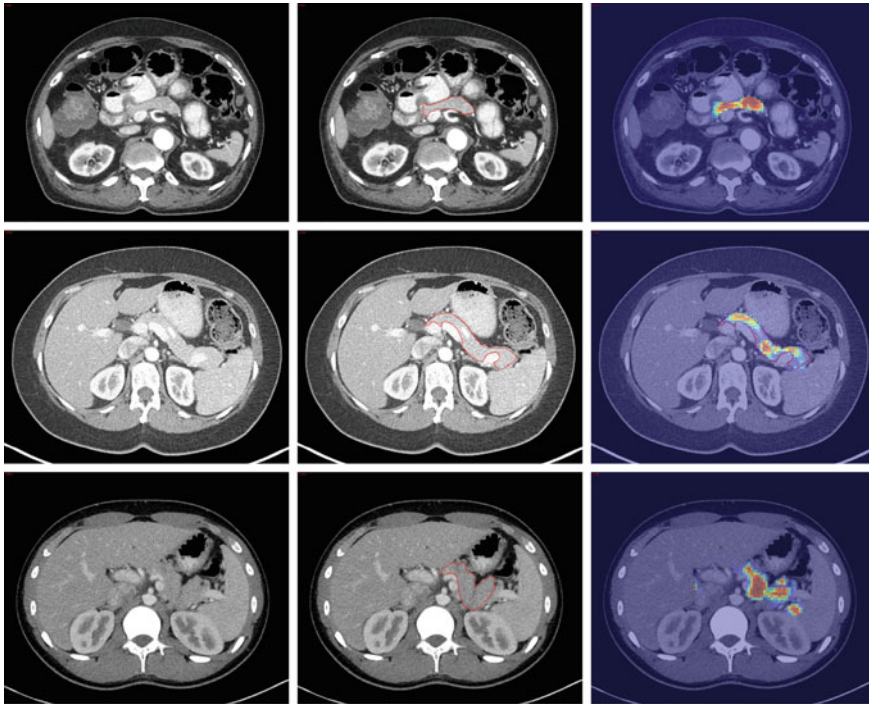
Fig. 16.10 Three examples of deep CNN-based image patch labeling probability response maps per row. *Red color* shows stronger pancreas class response and *blue* presents weaker response. From *Left*, *Center* to *Right* are the original CT image, CT image with annotated pancreas contour in *red*, and CNN response map overlaid CT image

Table 16.1 Examination of varying number of patient datasets using framework 1, in four metrics of Dice, JI, precision, and recall. Mean, standard deviation, lower and upper performance ranges are reported. Comparison of the presented framework 1 (F-1) versus framework 2 (F-2) in 80 patients is also presented

	N	SI (%)	JI (%)	Precision (%)	Recall (%)
F-1	40	60.4 ± 22.3	46.7 ± 22.8	55.6 ± 29.8	80.8 ± 21.2
F-1	60	64.9 ± 22.6	51.7 ± 22.6	70.3 ± 29.0	69.1 ± 25.7
F-1	80	68.8 ± 25.6	57.2 ± 25.4	71.5 ± 30.0	72.5 ± 27.2
F-2	80	70.7 ± 13.0	57.9 ± 13.6	71.6 ± 10.5	74.4 ± 15.1

13.0% in Dice; and 25.4–13.6% in JI). Note that F-1 has the similar standard deviation ranges with the previous methods [5, 7–10] and F-2 significantly improves upon all of them. From Figs. 16.1 and 16.6 it can be inferred that using the relative x-axis and y-axis positions as features aided in reducing the overall false negative rates. Based on Table 16.1, we observe that F-2 provides consistent performance improvements over F-1, which implies that CNN based dense patch labeling shows more
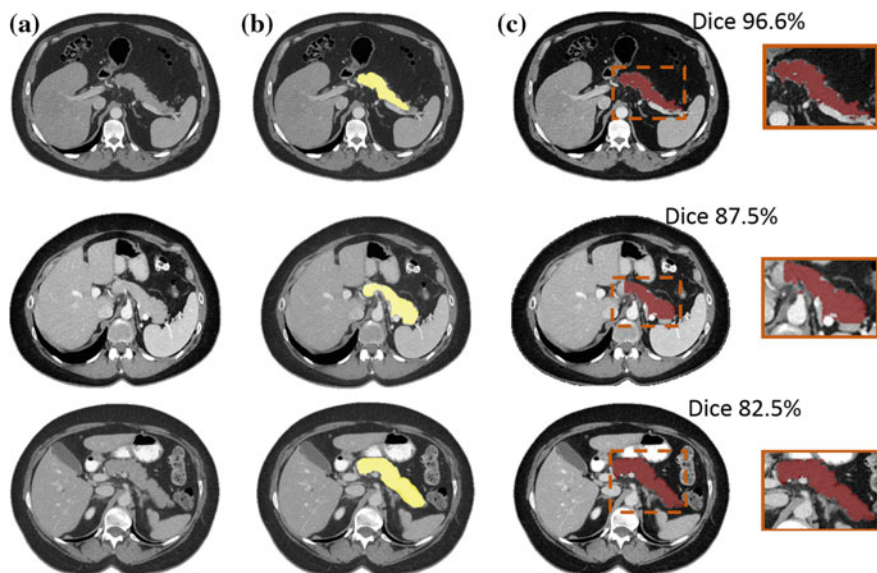
Fig. 16.11 Pancreas segmentation results with the computed Dice coefficients for one good (*Top Row*) and two fair (*Middle, Bottom Rows*) segmentation examples. Sample original CT slices for both patients are shown in (*Left Column*) and the corresponding ground-truth manual segmentation in (*Middle Column*) are in *yellow*. Final computed segmentation regions are shown in *red* in (*Right Column*) with Dice coefficients for the volume above each slice. The zoomed-in areas of the slice segmentation in the *orange boxes* are shown to the *right* of the image

promising results (Sect. 16.3.3) than the conventional had-crafted image features and random forest patch classification alone (Sect. 16.3.2). Figure 16.12 depicts an example patient where F-2 Dice score is improved by 18.6% over F-1 (from 63.9 to 82.5%). In this particular case, the close proximity of the stomach and duodenum to the pancreas head in particular proves challenging for F-1 without the CNN counterpart to distinguish. The surface-to-surface overlays illustrates how both frameworks compare to the ground-truth manual segmentation.

F-1 performs comparably to the state-of-the-art pancreas segmentation methods while F-2 slightly but consistently outperform others, even under sixfold cross validation (CV) instead of the "leave-one-patient-out" (LOO) used in [5–10]. Note that our results are not directly or strictly comparable with [5–10] since different datasets are used for evaluation. If under the same sixfold cross validation, our bottom-up segmentation method can significantly outperform an implemented version of "multi-atlas and label fusion" (MALF) based on [11, 12], on the pancreas segmentation dataset studied in this work. Details are provided later in this section. Table 16.2 reflects the comparison of Dice, JI, precision and recall results, between our methods of F-1, F-2 and other approaches, in multi-atlas registration and label fusion based multi-organ
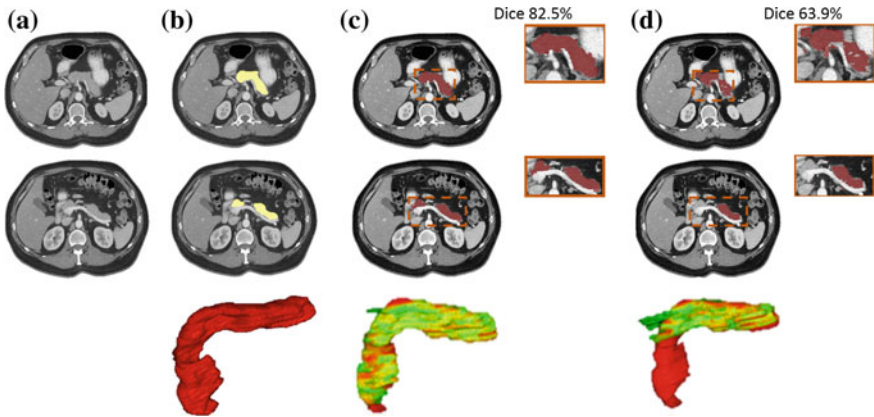
Fig. 16.12 Examples of pancreas segmentation results using F-1 and F-2 with the computed Dice coefficients for one patient. Original CT slices for the patient are shown in **Column a** and the corresponding ground-truth manual segmentation in **Column b** are in *yellow*. Final computed segmentation using F-2 and F-1 are shown in *red* in **Columns c, d** with Dice coefficients for the volume above first slice. The zoomed-in areas of the slice segmentation in the *orange boxes* are shown to the *right* of the images. Their surface-to-surface distance map overlaid on the ground-truth mask is demonstrated in **Columns c, d Bottom** and the corresponding ground-truth segmentation mask in **Column b Bottom** are in *red*. The *red color* illustrates higher difference and *green* for smaller distance

Table 16.2 Comparison of F-1 and F-2 in sixfold cross validation to the recent state-of-the-art methods [5–10] in LOO and our implementation of "multi-atlas and label fusion" (MALF) using publicly available C++ code bases [11, 12] under the same sixfold cross validation. The proposed bottom-up pancreas segmentation methods of F-1 and F-2 significantly outperform their MALF counterpart: $68.8 \pm 25.6\%$ (F-1), $70.7 \pm 13.0\%$ (F-2) versus $52.51 \pm 20.84\%$ in Dice coefficients (mean±std)

Reference	N	SI (%)	JI (%)	Precision (%)	Recall (%)
[5]	20	–	57.9	–	–
[6]	28	–	46.6	–	–
[7]	150	69.6 ± 16.7	55.5 ± 17.1	67.9 ± 18.2	74.1 ± 17.1
[7]	50	58.2 ± 20.0	43.5 ± 17.8	–	–
[9]	100	65.5	49.6	70.7	62.9
[10]	100	65.5 ± 18.6	–	–	–
[8]	100	69.1 ± 15.3	54.6	–	–
Framework 1	80	68.8 ± 25.6	57.2 ± 25.4	71.5 ± 30.0	72.5 ± 27.2
Framework 2	80	70.7 ± 13.0	57.9 ± 13.6	71.6 ± 10.5	74.4 ± 15.1
MALF	80	52.5 ± 20.8	38.1 ± 18.3	–	–

segmentation [6–10] and multi-phase single organ (i.e., pancreas) segmentation [5]. Previous numerical results are found from the publications [5–10]. We choose the best result out of different parameter configurations in [8].

We exploit two variations of pancreas segmentation in a perspective of bottom-up information propagation from image patches to (segments) superpixels. Both frameworks are carried out in a sixfold cross-validation (CV) manner. Our protocol is arguably harder than the "leave-one-out" (LOO) criterion in [5, 7–10] since less patient datasets are used in training and more separate patient scans for testing. In fact, [7] does demonstrate a notable performance drop from using 149 patients in training versus 49 patients under LOO, i.e., the mean Dice coefficients decreased from $69.6 \pm 16.7\%$ to $58.2 \pm 20.0\%$. This indicates that the multi-atlas fusion approaches [5–10] may actually achieve lower segmentation accuracies than reported, if under the sixfold cross-validation protocol. At 40 patients, our result using framework 1 is 2.2% better than the reported results by [7] using 50 patients (Dice coefficients of 60.4% vs. 58.2%). Comparing to the usage of $N-1$ patient datasets directly in the memory for multi-atlas registration methods, our learned models are more compactly encoded into a series of patch- and superpixel-level random forest classifiers and the CNN classifier for patch labeling. The computational efficiency also has been drastically improved in the order of 6–8 min per testing case (using a mix of Matlab and C implementation, $\sim$50% time for superpixel generation), compared to others requiring 10 hours or more. The segmentation framework (F-2) using deep patch labeling confidences is also more numerically stable, with no complete failure case and noticeable lower standard deviations.

Comparison to R-CNN and its variations [33, 39]: The conventional approach for classifying superpixels or image segments in computer vision is "bag-of-words" [40, 41]. "Bag-of-words" methods compute dense SIFT, HOG, and LBP image descriptors, embed these descriptors through various feature encoding schemes and pool the features inside each superpixel for classification. Both model complexity and computational expense [40, 41] are very high, comparing with ours (Sect. 16.3.4). Recently, a "Regional CNN" (R-CNN) [34, 42] method is proposed and shows substantial performance gains in PASCAL VOC object detection and semantic segmentation benchmarks [37], compared to previous "Bag-of-words" models. A simple R-CNN implementation on pancreas segmentation has been explored in our previous work [39] which reports evidently worse result (Dice coefficient $62.9 \pm 16.1\%$) than our F-2 framework (Dice $70.7 \pm 13.0\%$) that spatially pools the CNN patch classification confidences per superpixel. Note that R-CNN [34, 42] is not an "end-to-end" trainable deep learning system: R-CNN first uses the pretrained or fine-tuned CNNs as image feature extractors for superpixels and then the computed deep image features are classified by support vector machine models.

Our recent work [33] is an extended version of pancreas segmentation from the region-based convolutional neural networks (R-CNN) for semantic image segmentation [37, 42]. In [33], (1) we exploit multi-level deep convolutional networks which sample a set of bounding boxes covering each image superpixel at multiple spatial scales in a zoom-out fashion [43]; (2) the best performing model in [33] is a stacked R^2-ConvNet which operates in the joint space of CT intensities and the *Patch*-ConvNet dense probability maps, similar to F-2. With the above two method extensions, [33] reports the Dice coefficient of $71.8 \pm 10.7\%$ in fourfold cross val-

idation (which is slightly better than $70.7 \pm 13.0\%$ of F-2 using the same dataset). However, [33] cannot be directly trained and tested on the raw CT scans as in this work, due to the data high-imbalance issue between pancreas and non-pancreas superpixels. There are overwhelmingly more negative instances than positive ones if training the CNN models directly on all image superpixels from abdominal CT scans. Therefore, given an input abdomen CT, an initial set of superpixel regions is first generated or filtered by a coarse cascading process of operating the random forests based pancreas segmentation [44] (similar to F-1), at low or conservative classification thresholds. Over 96% original volumetric abdominal CT scan space has been rejected for the next step. For pancreas segmentation, these pre-labeled superpixels serve as regional candidates with high sensitivity (>97%) but low precision (generally called Candidate Generation or CG process). The resulting initial DSC is 27% on average. Then [33] evaluates several variations of CNNs for segmentation refinement (or pruning). F-2 performs comparably to the extended R-CNN version for pancreas segmentation [33] and is able to run without using F-1 to generate pre-selected superpixel candidates (which nevertheless is required by [33, 39]). As discussed above, we would argue that these hybrid approaches combining or integrating deep and non-deep learning components (like this work and [33, 34, 39, 42, 45]) will co-exist with the other fully "end-to-end" trainable CNN systems [46, 47] that may produce comparable or even inferior segmentation accuracy levels. For example, [45] is a two-staged method of deep CNN image labeling followed by fully connected Conditional Random Field (CRF) post-optimization [48], achieving 71.6% intersection-over-union value versus 62.2% in [47], on PASCAL VOC 2012 test set for semantic segmentation task [37].

Comparison to MALF (under sixfold CV): For the ease of comparison to the previously well studied "multi-atlas and label fusion" (MALF) approaches, we implement a MALF solution for pancreas segmentation using the publicly available C++ code bases [11, 12]. The performance evaluation criterion is the same **sixfold patient splits for cross validation**, not the "leave-one-patient-out" (LOO) in [5–10]. Specifically, each atlas in the training folds is registered to every target CT image in the testing fold, by the fast free-form deformation algorithm developed in NiftyReg [11]. Cubic B-Splines are used to deform a source image to optimize an objective function based on the normalized mutual information and a bending energy term. Grid spacing along three axes are set as 5 mm. The weight of the bending energy term is 0.005 and the normalized mutual information with 64 bins are used. The optimization is performed in three coarse-to-fine levels and the maximal number of iterations per level is 300. More details can be found in [11]. The registrations are used to warp the pancreas in the atlas set (66, or 67 atlases) to the target image. Nearest neighbor interpolation is employed since the labels are binary images. For each voxel in the target image, each atlas provided an opinion about the label. The probability of pancreas at any voxel x in the target U was determined by $\hat{L}(x) = \sum_{i=1}^{n} \omega_i(x)L_i(x)$ where $L_i(x)$ is the warped i-th pancreas atlas and $\omega_i(x)$ is a weight assigned to the i-th atlas at location x with $\sum_{i=1}^{n} \omega_i(x) = 1$; and n is the number of atlases. In our sixfold cross validation experiments $n = 66$ or 67. We adopt the joint label fusion algorithm [12], which estimates

voting weights $\omega_i(x)$ by simultaneously considering the pairwise atlas correlations and local image appearance similarities at x. More details about how to capture the probability that different atlases produce the same label error at location x via a formulation of dependency matrix can be found in [12]. The final binary pancreas segmentation label or map $L(x)$ in target can be computed by thresholding on $\hat{L}(x)$. The resulted MALF segmentation accuracy in Dice coefficients are $52.51 \pm 20.84\%$ in the range of [0, 80.56%]. This pancreas segmentation accuracy is noticeably lower than the mean Dice scores of 58.2–69.6% reported in [5–10] under the protocol of "leave-one-patient-out" (LOO) for MALF methods. This observation may indicate the performance deterioration of MALF from LOO (equivalent to 80-fold CV) to sixfold CV which is consistent with the finding that the segmentation accuracy drops from 69.6 to 58.2% when only 49 atlases are available instead of 149 [7].

Furthermore, we take about 33.5 days to fully conduct the sixfold MALF cross-validation experiments using a Windows server; whereas the proposed bottom-up superpixel cascade approach finishes in ∼9 h for 80 cases (6.7 min per patient scan on average). In summary, using the same dataset and under sixfold cross validation, our bottom-up segmentation method significantly outperforms its MALF counterpart: $70.7 \pm 13.0\%$ versus $52.51 \pm 20.84\%$ in Dice coefficients, while being approximately 90 times faster. Converting our Matlab/C++ implementation into pure C++ should expect further 2–3 times speed-up.

16.5   Conclusion and Discussion

In this chapter, we present a fully-automated bottom-up approach for pancreas segmentation in abdominal computed tomography (CT) scans. The proposed method is based on a hierarchical cascade of information propagation by classifying image patches at different resolutions and multi-channel feature information pooling at (segments) superpixels. Our algorithm flow is a sequential process of decomposing CT slice images as a set of disjoint boundary-preserving superpixels; computing pancreas class probability maps via dense patch labeling; classifying superpixels via aggregating both intensity and probability information to form image features that are fed into the cascaded random forests; and enforcing a simple spatial connectivity based post-processing. The dense image patch labeling can be realized by efficient random forest classifier on handcrafted image histogram, location and texture features; or deep convolutional neural network classification on larger image windows (i.e., with more spatial contexts).

The main component of our method is to classify superpixels into either pancreas or non-pancreas class. Cascaded random forest classifiers are formulated for this task and performed on the pooled superpixel statistical features from intensity values and supervisedly learned class probabilities (P^{RF} and/or P^{CNN}). The learned class probability maps (e.g., P^{RF} and P^{CNN}) are treated as the supervised semantic class image embeddings which can be implemented, via an open framework by various methods, to learn the per-pixel class probability response.

To overcome the low image boundary contrast issue in superpixel generation, which is however common in medical imaging, we suggest that efficient supervised edge learning techniques may be utilized to artificially "enhance" the strength of semantic object-level boundary curves in 2D or surface in 3D. For example, one of the future directions is to couple or integrate the structured random forests based edge detection [49] into a new image segmentation framework (MCG: Multiscale Combinatorial Grouping) [50] which permits a user-customized image gradient map. This new approach may be capable to generate image superpixels that can preserve even very weak semantic object boundaries well (in the image gradient sense) and subsequently prevent segmentation leakage.

Finally, voxel-level pancreas segmentation masks can be propagated from the stacked superpixel-level classifications and further improved by an efficient boundary refinement post-processing, such as the narrow-band level-set based curve/surface evolution [26], or the learned intensity model based graph-cut [7]. Further examination into the sub-connectivity processes for the pancreas segmentation framework that considers the spatial relationships of splenic, portal, and superior mesenteric veins with pancreas may be needed for future work.

References

1. Cuingnet R, Prevost R, Lesage D, Cohen L, Mory B, Ardon R (2012) Automatic detection and segmentation of kidneys in 3d CT images using random forests. In: MICCAI, pp 66–74
2. Mansoor A, Bagci U, Xu Z, Foster B, Olivier K, Elinoff J, Suffredini A, Udupa J, Mollura D (2014) A generic approach to pathological lung segmentation. IEEE Trans Med Imaging 33(12):2293–2310
3. Zheng Y, Barbu A, Georgescu B, Scheuering M, Comaniciu D (2008) Four-chamber heart modeling and automatic segmentation for 3d cardiac CT volumes using marginal space learning and steerable features. IEEE Trans Med Imaging 27(11):1668–1681
4. Ling H, Zhou S, Zheng Y, Georgescu B, Suehling M, Comaniciu D (2008) Hierarchical, learning-based automatic liver segmentation. In: IEEE conference on CVPR, pp 1–8
5. Shimizu A, Kimoto T, Kobatake H, Nawano S, Shinozaki K (2010) Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography. Int J Comput Assist Radiol Surg 5(1):85–98
6. Okada T, Linguraru M, Yoshida Y, Hor M, Summers R, Chen Y, Tomiyama N, Sato Y (2012) Abdominal multi-organ segmentation of CT images based on hierarchical spatial modeling of organ interrelations. In: Abdominal imaging - computational and clinical applications, pp 173–180
7. Wolz R, Chu C, Misawa K, Fujiwara M, Mori K, Rueckert D (2013) Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE Trans Med Imaging 32(7):1723–1730
8. Chu C, Oda M, Kitasaka T, Misawa K, Fujiwara M, Hayashi Y, Nimura Y, Rueckert D, Mori K (2013) Multi-organ segmentation based on spatially-divided probabilistic atlas from 3d abdominal CT images. In: MICCAI, vol 2, pp 165–172
9. Wolz R, Chu C, Misawa K, Mori K, Rueckert D (2012) Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases. In: MICCAI, vol 1, pp 10–17
10. Wang Z, Bhatia K, Glocker B, Marvao A, Dawes T, Misawa K, Mori K, Rueckert D (2014) Geodesic patch-based segmentation. In: MICCAI, vol 1, pp 666–673

11. Modat M, McClelland J, Ourselin S (2010) Lung registration using the niftyreg package. In: Medical image analysis for the clinic-a grand challenge, pp 33–42
12. Wang H, Suh J, Das S, Pluta J, Craige C, Yushkevich P (2012) Multi-atlas segmentation with joint label fusion. IEEE Trans Pattern Anal Mach Intell 35(3):611–623
13. Zografos V, Menze B, Tombari F (2015) Hierarchical multi-organ segmentation without registration in 3d abdominal CT images. In: MICCAI medical computer vision workshop
14. Lucchi A, Smith K, Achanta R, Knott G, Fua P (2012) Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features. IEEE Trans Med Imaging 31(2):474–486
15. Lu L, Wu D, Lay N, Liu D, Nogues I, Summers R (2016) Accurate 3d bone segmentation in challenging CT images: bottom-up parsing and contextualized optimization. In: IEEE conference on WACV, pp 1–10
16. Lu L, Barbu A, Wolf M, Liang J, Bogoni L, Salganicoff M, Comaniciu D (2008) Simultaneous detection and registration for ileo-cecal valve detection in 3d CT colonography. In: European conference on computer vision. Springer, Berlin, pp 465–478
17. Liu M, Lu L, Ye X, Yu S, Huang H (2011) Coarse-to-fine classification using parametric and nonparametric models for computer-aided diagnosis. In: 20th ACM conference on information and knowledge management
18. Lu L, Devarakota P, Vikal S, Wu D, Zheng Y, Wolf M (2013) Computer aided diagnosis using multilevel image features on large-scale evaluation. In: Medical computer vision-MICCAI. Springer, Berlin, pp 161–174
19. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S (2012) Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell 34(11):2274–2282
20. Neubert P, Protzel P (2012) Superpixel benchmark and comparison. In: Proceedings of the forum Bildverarbeitung, pp 1–12
21. Liu M, Tuzel O, Ramalingam S, Chellappa R (2011) Entropy rate superpixel segmentation. In: IEEE conference on CVPR, pp 2099–2104
22. Felzenszwalb P, Huttenlocher D (2004) Efficient graph-based image segmentation. Int J Comput Vis 59(2):167–181
23. Vincent L, Soille P (1991) Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans Pattern Anal Mach Intell 13(6):583–598
24. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 24(5):603–619
25. Cour T, Benezit F, Shi J (2005) Spectral segmentation with multiscale graph decomposition. In: IEEE proceedings on CVPR
26. Kohlberger T, Sofka M, Zhang J, Birkbeck N, Wetzl J, Kaftan J, Declerck J, Zhou S (2011) Automatic multi-organ segmentation using learning-based segmentation and level set optimization. In: MICCAI, vol 3, pp 338–345
27. Vedaldi A, Fulkerson B (2008) VLFeat: an open and portable library of computer vision algorithms. http://www.vlfeat.org/
28. Lowe D (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
29. Gilinsky A, Zelnik-Manor L (2013) Siftpack: a compact representation for efficient sift matching. In: IEEE conference on ICCV, pp 777–784
30. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
31. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
32. Gao M, Bagci U, Lu L, Wu A, Buty M, Shin H-C, Roth H, Papadakis GZ, Depeursinge A, Summers RM et al (2016) Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. Comput Methods Biomech Biomed Eng: Imaging Vis 1–6
33. Roth H, Lu L, Farag A, Shin H, Liu J, Turkbey E, Summers R (2015) Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation. In: MICCAI, pp 556–564

34. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE conference on CVPR, pp 580–587
35. Groeneveld R, Meeden G (1984) Measuring skewness and kurtosis. Stat 33:391–399
36. Viola P, Jones M (2004) Robust real-time face detection. Int J Comput Vis 57(2):137–154
37. Everingham M, Eslami S, Van Gool L, Williams C, Winn J, Zisserman A (2015) The pascal visual object classes challenge: a retrospective. Int J Comput Vis 111(1):98–136
38. Carreira J, Sminchisescu C (2012) CPMC: automatic object segmentation using constrained parametric min-cuts. IEEE Trans Pattern Anal Mach Intell 34(7):1312–1328
39. Roth H, Farag A, Lu L, Turkbey E, Liu J, Summers R (2015) Deep convolutional networks for pancreas segmentation in CT imaging. In: SPIE conference on medical imaging, pp 1–8
40. Carreira J, Caseiro R, Batista J, Sminchisescu C (2012) Semantic segmentation with second-order pooling. In: European conference on computer vision, pp 430–443
41. Chatfield K, Lempitsky V, Vedaldi A, Zisserman A (2011) The devil is in the details: an evaluation of recent feature encoding methods. In: BMVC, pp 1–12
42. Girshick R, Donahue J, Darrell T, Malik J (2015) Region-based convolutional networks for accurate object detection and semantic segmentation. IEEE Trans Pattern Anal Mach Intell, to appear
43. Mostajabi M, Yadollahpour P, Shakhnarovich G (2015) Feedforward semantic segmentation with zoom-out features. In: IEEE conference on CVPR, pp 3376–3385
44. Farag A, Lu L, Liu J, Turkbey E, Summers R (2010) A bottom-up approach for automatic pancreas segmentation abdominal CT scans. In: MICCAI abdominal imaging workshop
45. Chen L, Papandreou G, Kokkinos I, Murphy K, Yuille A (2015) Semantic image segmentation with deep convolutional nets and fully connected CRFs. In: International conference on learning representation
46. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: MICCAI, pp 234–241
47. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: IEEE conference on CVPR, pp 3431–3440
48. Krhenbhl P, Koltun V (2011) Efficient inference in fully connected CRFs with Gaussian edge potentials. In: NIPS, pp 109–117
49. Dollr P, Zitnick L (2015) Fast edge detection using structured forests. IEEE Trans Pattern Anal Mach Intell 37:1558–1570
50. Arbelez P, Pont-Tuset J, Barron J, Marqus F, Malik J (2014) Multiscale combinatorial grouping. In: IEEE conference on CVPR, pp 328–335