

Segmentación multi-atlas de imágenes médicas con selección de atlas inteligente y control de calidad automático

Lucas Mansilla¹, Enzo Ferrante¹

¹ *Sinc(i)*, Instituto de Investigaciones en Señales, Sistemas e Inteligencia Computacional,
Universidad Nacional del Litoral / CONICET

Ciudad Universitaria UNL, Ruta Nacional Nº 168, km 472.4,
FICH, 4to Piso (3000) Santa Fe – Argentina
{lmansilla, eferrante}@sinc.unl.edu.ar

Abstract. La segmentación multi-atlas (MAS por sus siglas en inglés) es una de las técnicas más utilizadas para la segmentación anatómica de imágenes médicas. Esta técnica utiliza un conjunto de atlas (imágenes previamente etiquetadas por expertos) para estimar la segmentación correspondiente a una nueva imagen, por medio de la registración y fusión de dichos atlas. Dado que los atlas de una población pueden variar significativamente entre sí, la selección de aquellos atlas más apropiados contribuye a mejorar su desempeño. En este trabajo, se presenta un estudio comparativo de segmentación basada en multi-atlas que utiliza distintos métodos de selección del atlas aleatorios y basados en similitud de imagen en el contexto de imágenes de radiografía torácica, para medir el impacto dicha selección en la precisión del algoritmo. Adicionalmente, para evaluar el desempeño del modelo en contextos en donde no se dispone de segmentaciones de referencia (también conocidas como *ground-truth*), se propone el uso del índice de Reverse Classification Accuracy (RCA) como estimador automático que posibilita el control de calidad de las segmentaciones en ausencia de segmentaciones de referencia. Los resultados obtenidos en dos grandes bases de datos demuestran que (i) la selección de atlas basada en similitud mejora significativamente los resultados al segmentar imágenes radiográficas y (ii) por medio de RCA es posible determinar un valor umbral de referencia que permite descartar automáticamente las segmentaciones de mala calidad en ausencia de *ground-truth*, tarea fundamental si se pretende avanzar en la utilización de este tipo de estrategias en el ámbito clínico real.

Keywords: segmentación basada en multi-atlas, selección del atlas, reverse classification accuracy, segmentación, imágenes médicas

1 Introducción

La segmentación de imágenes es un problema de gran interés para la comunidad médica dado que constituye un paso fundamental en la extracción de información útil para el diagnóstico, seguimiento y tratamiento de diversas patologías. En particular, la segmentación automática de estructuras anatómicas en imágenes de radiografía de tórax (tales como pulmones y corazón) reviste gran importancia para el diagnóstico asistido por computadora por medio de estas imágenes dado que facilita, entre otras cosas, la cuantificación de diversas propiedades morfológicas en dichas estructuras [11]. En la actualidad, una amplia gama de enfoques ha sido propuesta para resolver el problema de segmentación anatómica, siendo la segmentación multi-atlas [1] uno de los más adoptados debido a su robustez.

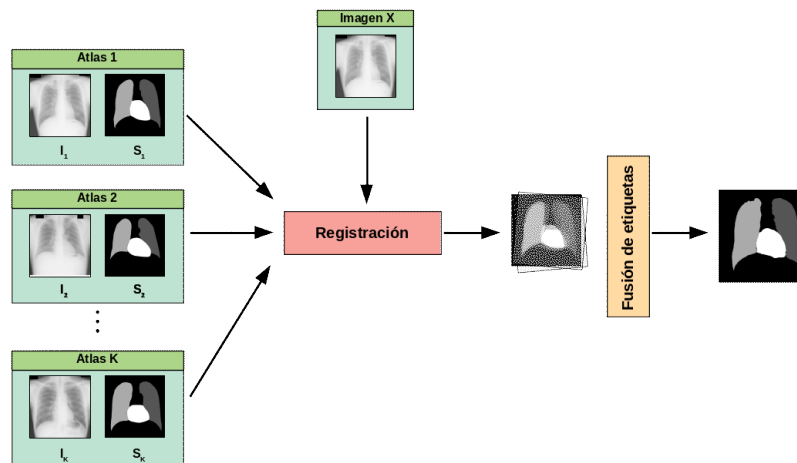


Fig. 1. Esquema del método de segmentación multi-atlas (MAS). La imagen de intensidad I_i de cada atlas se transforma para que resulte similar a la imagen a segmentar X . Esta transformación se aplica a la máscara de segmentación S_i de cada atlas. La segmentación final se obtiene mediante la fusión de las máscaras deformadas.

La segmentación multi-atlas (MAS) [1] utiliza un conjunto de atlas, esto es, pares formados por una imagen y su correspondiente máscara de segmentación, para obtener nuevas segmentaciones sobre imágenes que aún no han sido etiquetadas. Para lograrlo, la imagen de intensidad de cada atlas se transforma para que resulte similar a la imagen a segmentar. Esta transformación se aplica a la máscara de segmentación de cada atlas, para luego obtener la segmentación final mediante la combinación de las máscaras deformadas. Este último paso es conocido como fusión de etiquetas.

El proceso de deformar una imagen para que resulte similar a otra se conoce como registración deformable de imágenes [10]. En este contexto, la imagen de intensidad del atlas se denomina imagen móvil M y la imagen a segmentar se denomina imagen fija F . El resultado de registrar la imagen móvil M en la imagen fija F es una transformación T^* de M a F . La transformación óptima T^* es obtenida minimizando la siguiente función de energía:

$$T^* = \arg \min_T \Phi(M \circ T, F) + R(T), \quad (1)$$

donde Φ corresponde al criterio de (dis)similitud entre las dos imágenes M y F , la función R corresponde al término de regularización que impone restricciones geométricas a la solución, y T parametriza una transformación que mapea cada punto de la imagen M con F . Resolviendo este problema de minimización, es posible computar la transformación óptima T^* alineando dichas imágenes. Nótese que la transformación T^* puede ser lineal o no lineal. En este trabajo utilizaremos transformaciones no lineales deformables (representadas por un campo vectorial que indica las deformaciones a nivel pixel) dado que los métodos multi-atlas alcanzan mayor precisión cuando las imágenes son deformadas localmente antes de fusionar las etiquetas. En segmentación multi-atlas, esta transformación T^* es aplicada a la segmentación S de M en un proceso conocido como propagación de etiquetas, para obtener una segmentación deformada que representa una segmentación candidata para la imagen fija F [2].

La selección del atlas es un punto importante de todo algoritmo MAS. Inicialmente, se puede utilizar el total de los atlas disponibles en el dataset. Sin embargo, también es posible considerar un subconjunto del total disponible, dado que el proceso de registración suele ser costoso en términos computacionales. De esta forma, es posible mejorar la eficiencia computacional del algoritmo al tener una mejor cantidad de atlas que registrar. Otro punto a tomar en cuenta es que la segmentación final no depende sólo de una única registración sino de cada una de las registraciones de los atlas seleccionados, por lo que un proceso de selección de los atlas puede ayudar a mejorar la precisión de los resultados finales obtenidos, por ejemplo, descartando aquellos atlas que son muy diferentes a la imagen a segmentar y considerando sólo los que resulten más similares a esta [1, 2].

El nivel de precisión de un método de segmentación se suele medir evaluando la calidad de las segmentaciones resultantes. Tradicionalmente, esto es llevado a cabo usando diferentes métricas [3], en donde la segmentación predicha es comparada con una segmentación de referencia, también llamada *ground-truth* (GT), generada manualmente por un experto. La medida más popular en el contexto de la segmentación de imágenes médicas es el coeficiente de Dice-Sorensen (DSC) que

mide el solapamiento entre la segmentación estimada Y y la segmentación de referencia GT como:

$$DSC(Y, GT) = \frac{2|Y \cap GT|}{|Y| + |GT|} \quad (2)$$

Sin embargo, cuando se desea evaluar la performance de un algoritmo de segmentación sobre bases de datos sin ground-truth, no es posible aplicar estos métodos debido a la ausencia de máscaras de referencia. Este escenario es en realidad el más usual, dado que en la práctica cotidiana los métodos son desarrollados para segmentar imágenes que aún no han sido etiquetadas, y disminuir el trabajo de los expertos médicos evitando que realicen anotaciones manualmente. En este trabajo, adoptaremos la técnica Reverse Classification Accuracy (RCA) [4] para evaluar la calidad de la segmentación de una imagen individual en ausencia de GT, incorporando de esta forma control de calidad automático al proceso de segmentación. En este método, la segmentación predicha sirve de “ground truth” para entrenar un clasificador RCA que luego es evaluado con un conjunto de imágenes que sí disponen de GT. El mejor de los resultados de esta evaluación es el que se utiliza como medida de referencia para evaluar la calidad del método de segmentación.

Contribuciones de este trabajo: Este trabajo presenta dos contribuciones mayores: (i) en primer lugar, un estudio comparativo de la influencia de la técnica de selección del atlas en la calidad de las segmentaciones obtenidas en el contexto de imágenes de radiografía torácicas, evaluado en la base de datos pública JSRT [6] y alcanzando resultados del estado del arte; y (ii) la generación de máscaras de segmentación para un subconjunto del dataset público X-Chest [9] publicado recientemente (el cual no cuenta con segmentaciones manuales) incluyendo, para cada una, un estimador de la calidad de la segmentación obtenido mediante RCA que permite descartar segmentaciones erróneas.

2 Método

El método implementado para la segmentación y control de calidad automático de los resultados se compone de cuatro etapas que se describen detalladamente a continuación.

1. Selección inteligente de los atlas: El modelo propuesto de segmentación basada en multi-atlas incorpora un componente de selección de atlas para mejorar la calidad de las segmentaciones, y un componente RCA para el control automático de las mismas. Supongamos que se dispone de un conjunto de N atlas $\{(I_1, S_1), (I_2, S_2), \dots, (I_N, S_N)\}$ donde I_i corresponde a una imagen en escala de grises y S_i a su

correspondiente segmentación. De este conjunto, se seleccionan los K atlas $\{(I_1, S_1), (I_2, S_2), \dots, (I_K, S_K)\}$ más similares a la imagen que desea segmentar X , según una medida de similaridad de imagen de nida. Existen diversas medidas de similitud entre imágenes, algunas de ellas basadas en principios estadísticos como la información mutua (MI) [8] (considerando las distribuciones de intensidad en ambas imágenes como si fueran distribuciones de probabilidad y estimando la distribución conjunta) o algunas más simples como la distancia Euclidiana (D2) entre las imágenes I_i y X vectorizadas. En este trabajo, evaluaremos ambas medidas como método de selección de atlas y las compararemos con un método de selección aleatorio.

2. Registración de las imágenes: Cada una de las K imágenes I_i (seleccionadas en el paso anterior) es registrada con la imagen X de forma independiente. En dichos procesos de registración, las imágenes I_i son consideradas móviles mientras que la imagen X es considerada la imagen fija (ver Fig. 1).

Dicha registración es realizada minimizando la ecuación (1) de nida en la sección anterior. En este trabajo se utilizó el paquete SimpleElastix [7] que implementa algoritmos modulares de registración deformable. Como función de disimilitud Φ se utilizó una medida basada en la información mutua entre las imágenes. Adicionalmente, como función de regularización R se utilizó un término de penalización basado en la extensión del campo de deformación [12]. Al analizar el proceso de registración de dichas imágenes, se obtienen K segmentaciones deformadas $\{D_1, D_2, \dots, D_K\}$, todas ellas alineadas al espacio de coordenadas de la imagen X .

3. Fusión de etiquetas: Las segmentaciones obtenidas son combinadas utilizando un mecanismo de fusión de etiquetas resultando en una segmentación Y correspondiente a la imagen X . Uno de los mecanismos más simples utilizados para realizar fusión de etiquetas es el voto por mayoría. En ese caso, para cada pixel x de la imagen $X(x)$ se analiza la etiqueta otorgada por cada una de las K segmentaciones $\{D_1(x), D_2(x), \dots, D_K(x)\}$ y se asigna aquella que reciba la mayor cantidad de votos. Este mecanismo es sumamente simple, pero no incorpora al proceso de fusión de etiquetas información proveniente de los píxeles vecinos, resultando en ocasiones en segmentaciones espúreas alejadas de la realidad. Por esta razón, en este trabajo utilizaremos un método avanzado de fusión de etiquetas conocido *simultaneous truth and performance level estimation (STAPLE)* [5] que intenta maximizar el consenso entre las etiquetas imponiendo a la vez restricciones de homogeneidad entre píxeles vecinos.

4. Estimación de la calidad de la segmentación por RCA: La idea detrás del concepto de Reverse Classification Accuracy [4] es simple: en escenarios donde no contamos con segmentaciones de referencia en la imagen X para evaluar la calidad de la segmentación obtenida Y , utilizaremos el par (X, Y) de forma inversa, es decir, entrenaremos un algoritmo de segmentación usando sólo el par (X, Y) para segmentar

imágenes de las cuales sí tenemos segmentaciones de referencia. En el contexto de este trabajo, utilizaremos un algoritmo de segmentación basado en un único atlas como estimador RCA en lugar de múltiples atlas, es decir, que no poseerá etapa de fusión de etiquetas dado que contaremos con una única segmentación candidata. El par (X, Y) será utilizado como atlas único para segmentar las imágenes de las cuales poseemos segmentaciones de referencia. Luego, se evaluará la calidad de los resultados obtenidos, y tomaremos el máximo de dichas calidades como un estimador de la calidad real de la segmentación Y . La hipótesis sobre la que se sustenta este método es que si la imagen X fue segmentada correctamente con la máscara Y , entonces el par (X, Y) servirá como atlas para segmentar correctamente al menos una imagen del dataset con referencias (aquella que sea más parecida a la imagen X). La métrica a utilizar para cuantificar la calidad de la segmentación será el coeficiente DSC definido en la ecuación 2. Para una discusión completa sobre RCA ver [4].

3 Experimentos

Datasets. El modelo propuesto se evaluó en la tarea de segmentación de estructuras anatómicas en imágenes de radiografía de tórax. En primera instancia, se evaluó el desempeño del modelo bajo distintos métodos de selección del atlas, utilizando para ello la base de datos JSRT [6]. Esta es una base de datos de acceso público, compuesta por 247 imágenes de radiografías de tórax AP de 1024x1024 píxeles. Para cada imagen, se proveen segmentaciones manuales de pulmones y corazón, que serán utilizadas en este trabajo para evaluar el impacto de la estrategia de selección del atlas en los resultados obtenidos. Notar que en este caso se cuenta con segmentaciones de referencia para todas las imágenes, razón por la cual el coeficiente DSC puede ser calculado directamente sin necesidad de estimarlo por medio de RCA.

En segunda instancia, se evaluó el desempeño del modelo con RCA utilizando una muestra aleatoria de 500 imágenes de radiografías de tórax AP sin segmentaciones de referencia, extraída del dataset X-Chest [9] recientemente publicado. Dicho dataset no cuenta con segmentaciones anatómicas a nivel pixel, por lo que se utilizó el método RCA para estimar el coeficiente DSC y poder realizar un control automático de calidad que permita descartar segmentaciones defectuosas. En ambos casos, todas las imágenes utilizadas fueron reducidas a un tamaño de 256x256 píxeles para disminuir el costo computacional del método, y las segmentaciones obtenidas fueron posteriormente resampleadas a la resolución original de las imágenes.

Estudio comparativo de métodos para selección de atlas. Del conjunto de 247 imágenes con segmentaciones de referencia que componen la base de datos JSRT, 50 fueron utilizadas para evaluar el modelo bajo distintos métodos de selección del atlas, utilizando las restantes 197 como base de datos a partir de las cuales se seleccionaron

los atlas utilizados para segmentar cada imagen. De esta manera, se corrieron 3 instancias del algoritmo, cada una con un método de selección diferente, utilizando $K=5$ atlas para segmentar cada imagen. La Tabla 1 muestra el DSC promedio en cada estructura anatómica para selección por información mutua (MI), selección por mínima distancia Euclidiana al cuadrado (D2) y selección aleatoria (Random), alcanzando resultados comparables al estado del arte para JSRT reportados en [11].

Tabla 1. Valores de DSC promedio por etiqueta utilizando información mutua (MI), distancia Euclidiana al cuadrado (D2) y selección aleatoria (Random) de atlas. En cada estructura, el DSC final se obtuvo promediando entre los DSC calculados para esa estructura en particular en las 50 imágenes de test.

Método	Pulmón derecho	Pulmón izquierdo	Corazón
MI	0.940 (0.020)	0.950 (0.021)	0.882 (0.050)
D2	0.932 (0.034)	0.944 (0.030)	0.862 (0.065)
Random	0.920 (0.031)	0.929 (0.029)	0.826 (0.109)

Para una mejor visualización, en la Fig. 2 se muestra la distribución de valores de DSC en cada estructura utilizando boxplots. El modelo con información mutua es el que logró mejores medidas de DSC para todas las estructuras anatómicas.

Control de calidad automático con RCA. Dado que el modelo de segmentación MAS con selección de atlas candidatos por información mutua fue el que presentó mejor rendimiento, éste fue utilizado para segmentar las 500 imágenes del dataset X-Chest, utilizando RCA para estimar los valores de DSC, considerando igualmente $K=5$ atlas candidatos por cada imagen a segmentar. Cabe destacar que los atlas candidatos utilizados para segmentar las imágenes de X-Chest fueron seleccionados a partir de las 197 imágenes del dataset JSRT con anotaciones de referencia. En la Fig. 3 se muestra la distribución de valores de DSC estimados por medio de RCA para el dataset sin anotaciones de referencia. Para incorporar el control de calidad automático, se calculó el valor de DSC mínimo (DSC_{\min}) necesario que debe tener una imagen segmentada para no ser descartada. Siendo 0.88 y 0.04, la media y la desviación estándar de la distribución de valores de DSC estimados con RCA, se estableció $DSC_{\min} = 0.88 - 0.04 = 0.84$. Así, una imagen segmentada con un valor de DSC inferior a 0.84 será descartada automáticamente.

En la Fig. 4 se muestran algunas segmentaciones obtenidas junto con el DSC calculado por RCA. Se puede apreciar visualmente que el modelo es capaz de producir resultados de segmentación muy precisos, como es el caso de los dos ejemplos de la parte superior. Los dos restantes de la parte inferior son ejemplos en donde el modelo no se comporta adecuadamente, en este caso, estas segmentaciones resultarían descartadas automáticamente por tener un valor DSC menor a DSC_{\min} .

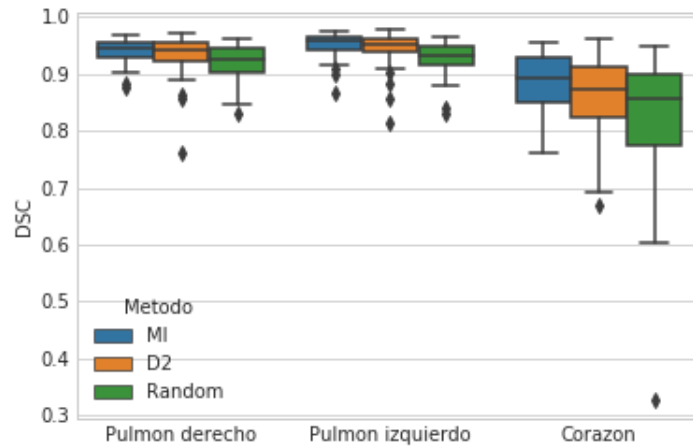


Fig. 2. Valores DSC por etiqueta utilizando información mutua (MI), distancia Euclidiana al cuadrado (D2) y selección aleatoria de atlas. Para medir la significancia estadística de las diferencias entre las distribuciones se utilizó el test de Wilcoxon por cada posible combinación de métodos, que arrojó p-value < 0.05 para todos los casos, siendo la única excepción el par MI-D2 para la etiqueta corazón.

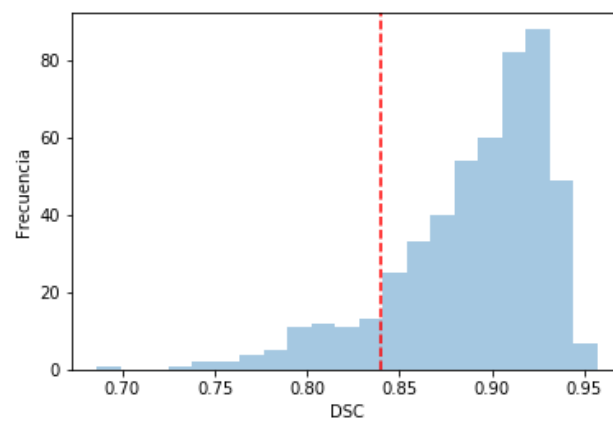


Fig. 3. Distribución de valores de DSC estimados con RCA. La línea vertical roja vertical representa el mínimo valor de DSC ($DSC_{\min}=0.84$) necesario que debe tener una imagen segmentada para no ser descartada.

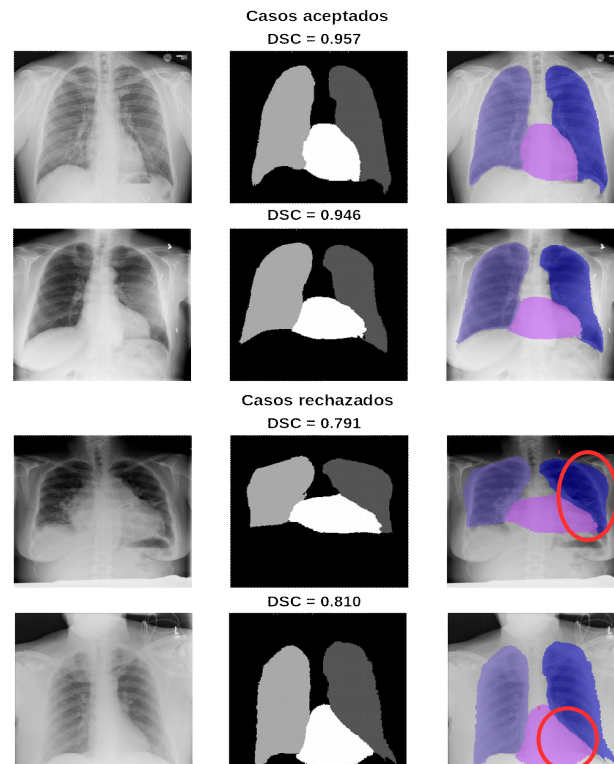


Fig. 4. Ejemplos de segmentaciones obtenidas junto con su correspondiente DSC estimado por RCA. (Izquierda) Ejemplos de imágenes a segmentar. (Centro) Segmentaciones producidas por multi-atlas con selección por MI. (Derecha) Se muestran los pares alineados. Puede observarse que en los casos rechazados por el control automático de calidad, las segmentaciones resultantes poseen serios errores que han sido resaltados en rojo.

4 Conclusiones y trabajos futuros

En este trabajo se presentó un método de segmentación basado en multi-atlas, que incorpora un componente de selección de atlas candidatos basado en similitud y utiliza RCA para medir la calidad de la segmentación en contextos donde no se dispone de segmentaciones de referencia. Los resultados obtenidos muestran que, cuando se utilizan métodos basados en similaridad de imagen para seleccionar atlas candidatos en el contexto de imágenes radiográficas de torax, la precisión del modelo aumenta significativamente. Para evaluar el desempeño del mejor modelo de segmentación obtenido cuando no se dispone de información de referencia, se incorporó la técnica RCA para estimar la calidad de la segmentación en ausencia de

GT. A partir de los valores de DSC estimados por RCA, se pudo definir un valor mínimo (DSC_{\min}) para efectuar un control de calidad de las segmentaciones de forma automática y detectar segmentaciones defectuosas alejadas de la realidad.

El uso combinado de RCA y segmentación multi-atlas abre las puertas al desarrollo de métodos que utilicen dichas segmentaciones (junto con las estimaciones de calidad provistas) como insumo para, por ejemplo, clasificar dichas imágenes como sanas o patológicas. El dataset utilizado X-Chest provee, para cada imagen, el diagnóstico médico indicando si la misma corresponde a un paciente sano o posee alguna patología como cardiomegalia o pneumo-torax, entre otras. En el futuro, se pretende utilizar la información provista por las segmentaciones obtenidas en conjunto con su indicador de calidad y las imágenes originales para intentar mejorar la tasa de acierto de los métodos existentes para el diagnóstico automático de estas enfermedades [9].

Referencias

1. Iglesias, J.E., Sabuncu, M.R.: Multi-atlas segmentation of biomedical images: a survey. *Medical Image Analysis*, 24(1), 205-219 (2015)
2. Agarwal, M.: Multi-Atlas Based Image Segmentation Applied on Pulmonary CT Scan Data (2018)
3. Crum, W. R., Camara, O., Hill, D. L.: Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE TMI*, 25(11), 1451 -1461 (2006)
4. Valindria, V. et al.: Reverse Classification Accuracy: Predicting Segmentation Performance in the Absence of Ground Truth. *IEEE TMI*, 36(8), 1597-1606 (2017)
5. Warfield, S. et al.: Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation. *IEEE TMI*, 23(7), 903-921 (2004)
6. Shiraishi, J. et al.: Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *AJR* 174; 71-74 (2000)
7. Marstal, K., Berendsen, F., Staring, M., Klein, S.: SimpleElastix: a user-friendly, multi-lingual library for medical image registration. In: *IEEE CVPR* (2016)
8. Viola P., Wells, W.M.: Alignment by Maximization of Mutual Information *International Journal of Computer Vision*, 24(2) 137-154 (1997)
9. Wang, X. et al.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017)
10. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging*, 32(7), 1153-1190 (2013)
11. Van Ginneken, B. et al.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *MedIA*, 10(1), 19-40 (2006)
12. Rueckert, D. et al.: Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Med. Imaging* 18(8), 712-721 (1999)