



Automatic Shadow Detection in 2D Ultrasound Images

Qingjie Meng¹(✉), Christian Baumgartner², Matthew Sinclair¹,
James Housden³, Martin Rajchl¹, Alberto Gomez³, Benjamin Hou¹,
Nicolas Toussaint³, Veronika Zimmer³, Jeremy Tan¹, Jacqueline Matthew³,
Daniel Rueckert¹, Julia Schnabel³, and Bernhard Kainz¹

¹ Biomedical Image Analysis Group, Imperial College London, London, UK
q.meng16@imperial.ac.uk

² Computer Vision Lab, ETH Zürich, Zürich, Switzerland

³ School of Biomedical Engineering and Imaging Sciences,
Kings College London, London, UK

Abstract. Automatically detecting acoustic shadows is of great importance for automatic 2D ultrasound analysis ranging from anatomy segmentation to landmark detection. However, variation in shape and similarity in intensity to other structures make shadow detection a very challenging task. In this paper, we propose an automatic shadow detection method to generate a pixel-wise, shadow-focused confidence map from weakly labelled, anatomically-focused images. Our method: (1) initializes potential shadow areas based on a classification task. (2) extends potential shadow areas using a GAN model. (3) adds intensity information to generate the final confidence map using a distance matrix. The proposed method accurately highlights the shadow areas in 2D ultrasound datasets comprising standard view planes as acquired during fetal screening. Moreover, the proposed method outperforms the state-of-the-art quantitatively and improves failure cases for automatic biometric measurement.

1 Introduction

2D Ultrasound (US) imaging is a popular medical imaging modality based on reflection and scattering of high frequency sound in tissue, well known for its portability, low cost, and high temporal resolution. However, this modality is inherently prone to artefacts in clinical practice due to low energies used and the physical nature of sound waves propagation in tissue. Artefacts such as noise, distortions and acoustic shadows are unavoidable, and have a significant impact on the achievable image quality. Noise can be handled through better hardware and advanced image reconstruction algorithms [7], while distortions can be tackled by operator training and knowledge of the underlying anatomy [15]. However, acoustic shadows are more challenging to resolve.

Acoustic shadows are caused by sound-opaque occluders, which can potentially conceal vital anatomical information. Shadow regions have low signal

intensity with very high acoustic impedance differences at the boundaries. Sonographers are trained to avoid acoustic shadows by using real-time acquisition devices. Shadows are either avoided by moving to a more preferable viewing direction or, if no shadow-free viewing direction can be found, a mental map is compounded with iterative acquisitions from different orientations. Although acoustic shadows may be useful for practitioners to determine the anatomical properties of occluders, images containing strong shadows can be problematic for automatic real-time image analysis methods which, such as; provide directional guidance; perform biometric measurements; or automatically evaluate biomarkers, etc. Therefore shadow-aware US image analysis would be beneficial for many of these applications, as well as clinical practice.

Contribution: (1) We propose a novel method that uses weak annotations (shadow/shadow-free images) to generate an anatomically agnostic shadow confidence map in 2D ultrasound images; (2) The proposed method achieves accurate shadow detection visually and quantitatively for different fetal anatomies; (3) To our knowledge, this is the first shadow detection model for ultrasound images that generates a dense, shadow-focused confidence map; (4) The proposed shadow detection method can be used in real-time automatic US image analysis, such as anatomical segmentation and registration. In our experiments, the obtained shadow confidence map greatly improves segmentation performance of failure cases in automatic biometric measurement.

Related Work: US artefacts have been well studied in clinical literature, e.g. [5, 13] provide an overview. However, anatomically agnostic acoustic shadow detection has rarely been the focus within the medical image analysis community. [10] developed a shadow detection method based on geometrical modelling of the US B-Mode cone with statistical tests. This is an anatomical-specific technique designed to detect only a subset of ‘deep’ acoustic shadows, which has shown improvements in 3D reconstruction/registration/tracking. [11] proposed a more general solution using the Random Walks (RW) algorithm for US attenuation estimation and shadow detection. In their work, ultrasound confidence maps are obtained to classify the reliability of US intensity information, and thus, to detect regions of acoustic shadow. Their approach yields good results for 3D US compounding but is sensitive to US transducer settings. [12] further extended the RW method to generate distribution-based confidence maps for a specific Radio Frequency (RF) US data. Other applications, such as [4, 6], use acoustic shadow detection as additional information in their pipeline. In both works, acoustic shadow detection functions as task-specific components, and is mainly based on image intensity features and the special anatomical constraints.

Advances in weakly supervised deep learning methods have drastically improved fully automatic semantic real-time image understanding [14, 17, 21]. However, most of these methods require pixel-wise labels for the training data, which is infeasible for acoustic shadows.

Unsupervised deep learning methods, showing visual attribution of different classes, have recently been developed in the context of Alzheimer’s disease classification from MRI brain scans [3].

Inspired by these works, we develop a method to identify potential shadow areas based on supervised classification of weakly labelled, anatomically-focused US images, and further extend the detection of potential shadow areas using the visual attribution from an unsupervised model. We then combine intensity features, extracted by a graph-cut model, with potential shadow areas to provide a pixel-wise, shadow-focused confidence map. The overview of the proposed method is shown in Fig. 1.

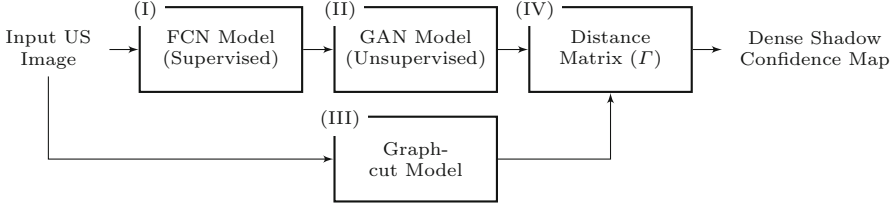


Fig. 1. Pipeline of the proposed method. (I) Identify potential shadow areas by a FCN model; (II) Extend obtained potential shadow areas using a GAN model; (III) Graph-cut is used to extract intensity features; (IV) The proposed distance matrix is designed to generate dense shadow confidence map from potential shadow areas and intensity features.

2 Method

Figure 2 shows a detailed inference flowchart over our method, which consists of four steps: (I) and (II) are used to highlight potential shadow areas, while step (III) selects coarse shadow areas based on intensity information. (IV) combines detection results from (II) and (III) to achieve the final shadow confidence map.

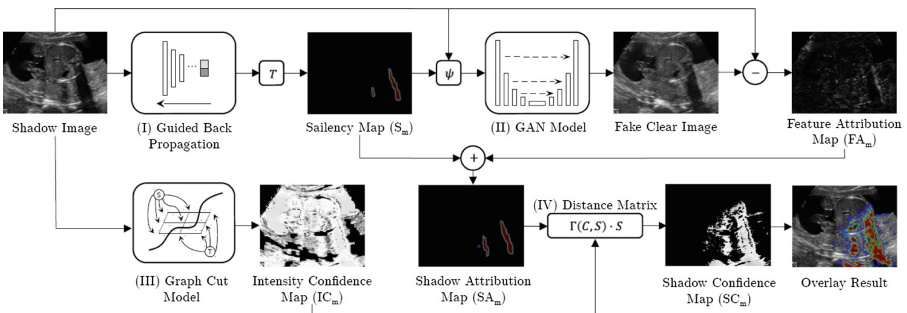


Fig. 2. Inference of our anatomy agnostic shadow detection approach.

(I) Saliency Map Generation: Saliency maps are generated by finding discriminative features from a trained classifier, using a gradient based back-propagation method, and thus, highlight distinct areas among different classes. Based on this property, it is a naïve approach to use saliency maps generated by shadow/shadow-free classifier for shadow detection.

We use a Fully Convolutional Neural-Network (FCN) to discern images containing shadows from shadow-free images. Here, we denote the has-shadow class with label $l = 1$ and the shadow-free class with label $l = 0$. Image set $X = \{x_1, x_2, \dots, x_K\}$ and their corresponding labels $L = \{l_1, l_2, \dots, l_K\}$ s.t. $l_i \in \{0, 1\}$ are used to train the FCN. The classifier provides predictions $p(x_i|l = 1)$ for image x_i during testing. We build the classifier model using SonoNet-32 [2], as it has shown promising results for 2D ultrasound fetal standard view classification. The training of the classifier is shown in Fig. 3.

Based on the trained shadow/shadow-free classifier, corresponding saliency maps $S_m = [s_{m1}, s_{m2}, \dots, s_{mN}]$ are generated by guided back-propagation [19] for N testing samples. Shadows typically have features such as directional occlusion with relatively low intensity. These features, highlighted in S_m , are potential shadow candidates on a per-pixel basis.

However, by using gradient based back-propagation, saliency maps may ignore some areas which are evidence of a class but may have no ultimate effect on the classification result. In the shadow detection task, obtained saliency maps focus mainly on the edge of shadow areas but may ignore the homogeneous centre of shadow areas.

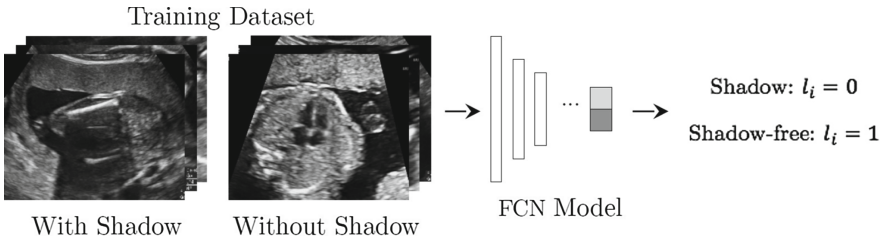


Fig. 3. Training FCN model for saliency map (S_m) generation

(II) Potential Shadow Areas Detection: Saliency maps heavily favour edges of the largest shadow region, especially when the image has multiple shadows, because these areas are the main difference between shadow and shadow-free images. In order to detect more shadows and inspired by VA-GAN [3], we develop a GAN model (shown in Fig. 4) that utilizes S_m to generate a Shadow Attribution Map (SA_m). S_m is used to inpaint the corresponding shadow image before passing the shadow image into the GAN model, so that the GAN model is forced to focus on other distinct areas between shadow and shadow-free images. Compared to S_m alone, this GAN model allows detection of more edges of relatively weak shadow areas as well as central areas of shadows.

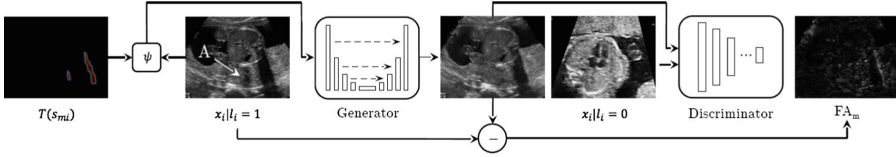


Fig. 4. Training GAN model for Feature Attribution map (FA_m) Generation.

The generator of the GAN model, G , produces a fake clear image from a shadow image x_i that has been inpainted with a binary mask of its corresponding saliency map. G has a U-Net structure with all its convolution layers being replaced by residual-units [9]. We optimize G by the Wasserstein distance [1], as it simplifies the optimization process and makes training more stable. The discriminator of the GAN model, D , is used to discern fake clear images from real clear images, and is trained with unpaired data. In the proposed method, the discriminator is a FCN without dense layers.

The inpainting function, used for the GAN input, is defined as $\psi := \psi(x_i | l_i = 1, T(s_{m_i}))$. Here, $T_b^a(\cdot)$ produces a pixel-wise binary mask to identify pixels that lie in the top a and bottom b percentile of the input's intensity histogram distribution. In our experiments, we take the 2nd and 98th percentile respectively of the saliency map, s.t. $T_2^{98}(s_{m_i}) = \{0 : P_2 \leq s_{m_i} \leq P_{98}, 1 : \text{otherwise}\}$. ψ then replaces pixels in $x_i(T_2^{98}(s_{m_i}) = 1)$ with the mean intensity value of $x_i(T_2^{98}(s_{m_i}) = 0)$. The generator therefore focuses on more ambiguous shadow areas, as well as the central areas of shadows, to generate the fake clear image.

The overall cost function (shown in Eq. 1) consists of the GAN model loss $\mathcal{L}_{GAN}(G, D)$, a L1-loss \mathcal{L}_1 and a L2-loss \mathcal{L}_2 . The $\mathcal{L}_{GAN}(G, D)$ is defined in Eq. 2. \mathcal{L}_1 is defined as in Eq. 3 to guarantee small changes in the output, while \mathcal{L}_2 is defined as Eq. 4 to encourage changes to happen only in potential shadow areas.

$$\mathcal{L} = \mathcal{L}_{GAN}(G, D) + \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_2 \quad (1)$$

$$\mathcal{L}_{GAN}(G, D) = \mathbf{E}_{\psi(\cdot) \sim p(\psi(\cdot) | l=0)}[D(x_i)] - \mathbf{E}_{\psi(\cdot) \sim p(\psi(\cdot) | l=1)}[D(G(\psi(\cdot)))] \quad (2)$$

$$\mathcal{L}_1 = \|G(\psi(\cdot)) - \psi(\cdot)\|_1 \quad (3)$$

$$\mathcal{L}_2 = \|G(\psi(\cdot)_B) - \psi(\cdot)_B\|_2 \quad (4)$$

We train the networks using the optimisation method from [8] and set the gradient penalty as 10. The parameters for the optimiser are $\beta_1 = 0$, $\beta_2 = 0.9$, with the learning rate 10^{-3} . In the first 30 iterations and every hundredth iteration, the discriminator updates 100 times for every update of the generator. In other iterations, the discriminator updates five times for every single update of the generator. We set the weights of the combined loss function to $\lambda_1 = 0$, $\lambda_2 = 0.1$ for the first 20 epochs and $\lambda_1 = 10^4$, $\lambda_2 = 0$ for the remaining epochs.

The Feature Attribution map, FA_m , defined in Eq. 5, is obtained by subtracting the generated fake clear image from the original shadow image. The Shadow Attribution map is then $SA_m = FA_m + S_m$.

$$FA_m = |G(\psi(x_i|l_i = 1, T(s_{m_i}))) - x_i| \quad (5)$$

(III) Graph Cut Model: Another feature of shadows is their relatively low intensity. To integrate this feature, we build a graph cut model using intensity information as weights to connect each pixel in the image to shadow class and background class. After using the Min-Cut/Max-Flow algorithm [20] to cut the graph, the model shows pixels belonging to the shadow class. The weights that connect pixels to the shadow class give an intensity saliency map IC_m .

Since shadow ground truth is not available for every image, we randomly select ten shadow images from training data for manual segmentation to compute the shadow mean intensity I_S . Background mean intensity I_B is computed by thresholding these ten images using the top 80th percentile.

For a pixel x_{ij} with intensity I_{ij} , the score of being a shadow pixel F_{ij} is given by Eq. 6 while the score of being a background pixel B_{ij} is given by Eq. 7. The weight from x_{ij} to source (shadow class) is set as $W_{F_{ij}} = \frac{F_{ij}}{F_{ij} + B_{ij}}$ and the weight from x_{ij} to sink (background) is $W_{B_{ij}} = \frac{B_{ij}}{F_{ij} + B_{ij}}$. We use a 4-connected neighbourhood to set weights between pixels and all the weights between neighbourhood pixels are set to 0.5.

$$F_{ij} = -\frac{|I_{ij} - I_S|}{|I_{ij} - I_S| + |I_{ij} - I_B|} \quad (6)$$

$$B_{ij} = -\frac{|I_{ij} - I_B|}{|I_{ij} - I_S| + |I_{ij} - I_B|} \quad (7)$$

(IV) Distance Matrix: Since the intensity distribution of shadow areas are homogeneous, potential shadow areas detected in SA_m from (II) are mainly edges of shadows. Meanwhile, IC_m from (III) shows all pixels with a similar intensity to shadow areas. In this step, we propose a distance matrix \mathbf{D} combining IC_m with SA_m to produce a Shadow Confidence Map (SC_m). In SC_m , pixels with a similar intensity to shadow areas and spatially closer to potential shadow areas achieves higher confidence of being part of shadow areas.

$$\Gamma(IC_m, SA_m) = 1 - \frac{Dis}{\max(Dis)} \quad (8)$$

$$SC_m = \Gamma(IC_m, SA_m) \cdot IC_m \quad (9)$$

The distance matrix is defined in Eq. 8. Dis is the set of the spatial distances that each pixel $IC_{m_{ij}}$ to potential shadow areas in SA_m . Each element Dis_{ij} in Dis refers to the smallest distance of $IC_{m_{ij}}$ to all connected components in SA_m . SC_m is obtained by multiplying the distance matrix Γ to IC_m (shown in Eq. 9) which leads to pixels with similar shadow area intensity and closer to the potential shadow areas achieve a higher score in SC_m .

3 Evaluation and Results

US Image Data: The data set used in our experiments consists of $\sim 8.5k$ 2D fetal ultrasound images sampled from 14 different anatomical standard plane locations as they are defined in the UK FASP handbook [16]. These images have been sampled from 2694 2D ultrasound examinations from volunteers with gestational ages between 18–22 weeks. Eight different ultrasound systems of identical make and model (GE Voluson E8) were used for the acquisitions. The images have been classified by expert observers as containing strong shadow, being clear, or being corrupted, e.g. lacking acoustic impedance gel. Corrupted images (<3%) have been excluded.

3448 shadow images and 3842 clear images have been randomly selected for data set **A**, which is used for training. The remaining 491 shadow images and 502 clear images are used for validation. Data set **B**, a subset from the 491 shadow validation images, comprises of 48 randomly selected non-brain images, where shadows have been manually segmented to provide ground truth.

An additional data set **C**, which has no overlap with the $\sim 8.5k$ fetal images, comprises of 643 fetal brain images. The entire data set C has been used for validation and shadows in this data set have been coarsely segmented by bio-engineering students.

We apply image flipping as data augmentation. Our models are trained on a Nvidia Titan X GPU with 12 GB of memory.

Table 1. Threshold ranges and DICE scores of different shadow detection methods: RW [11] vs. intermediate results from our approach and the final shadow confidence map.

	RW	S_m	FA_m	SA_m	SC_m
Dataset B	$T_3^{100}(S_m)$	$T_1^{99}(S_m)$	$T_1^{85}(FA_m)$	$T_1^{96}(SA_m)$	$T_0^{80}(SC_m)$
	0.06	0.25	0.06	0.27	0.55
Dataset C	$T_3^{100}(S_m)$	$T_1^{99}(S_m)$	$T_1^{80}(FA_m)$	$T_0^{90}(SA_m)$	$T_0^{70}(SC_m)$
	0.11	0.28	0.08	0.31	0.36

Experiment Results: The classification accuracy of the FCN classifier on the validation data set C is 94%. The FCN classifier’s saliency maps are shown in Fig. 5 column (b) for three examples from data set B and C.

To provide quantitative evaluation (Table 1), we chose the percentile range used by T for SC_m as well as other intermediate maps (S_m , FA_m , SA_m). These percentile ranges for different maps are chosen heuristically through experiments on validation data set B and C, so that these thresholded segmentation of data set B and C contains the most shadow areas and the least noise. We compare these thresholded segmentation with manual segmentation in data set B and C using the DICE score. Additionally, we compare the thresholded versions of the confidence map derived from the RW method [11]. The parameters for RW

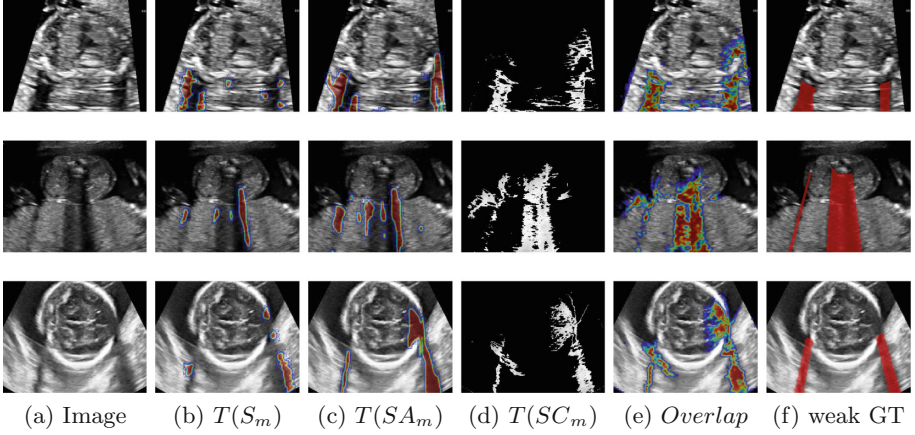


Fig. 5. Rows 1–3 show examples for shadow detection; Right Ventricular Outflow Tract (top), Kidney (middle), and an axial view through the brain (bottom). The key steps from Fig. 2 are illustrated from (a) the input image to (f) the coarse ground truth (GT) from manual segmentation.

in our experiments are: $\alpha = 1$; $\beta = 90$; $\gamma = 0.3$, which reach the highest DICE score on our validation data sets. Qualitative results are shown in Fig. 5. The GAN model in our approach is essential as it picks up less prominent shadows as shown in Fig. 6.

Application: We integrate SC_m as an additional channel in a clinical system that automatically measures cranial and abdominal circumferences [18]. This system is based on FCNs and works well for images without shadows but fails

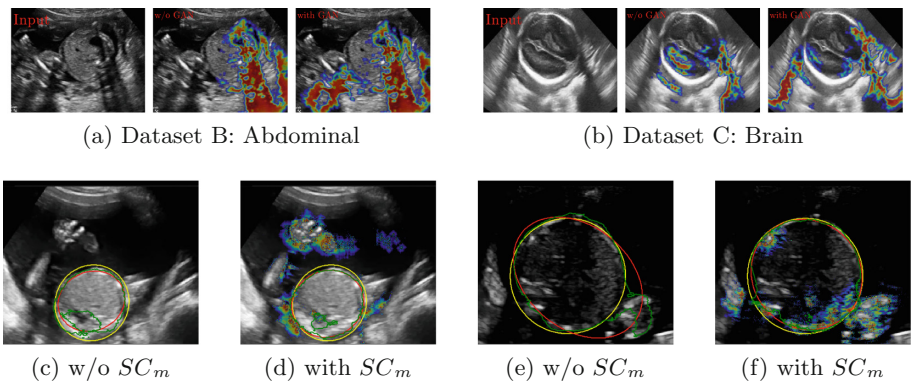


Fig. 6. (a–b) Two examples for the importance of the GAN model (input image – w/o GAN – with GAN). (c–f) Improving automatic biometric measurements through applying SC_m as additional channel to a FCN [18] (yellow = GT, red = prediction, green = segmentation boundary). (Color figure online)

for about 5–10% of abdominal test images which show strong shadows. By adding SC_m as an additional input channel, segmentation performance is boosted by up to 10% for individual failure cases, when measuring the DICE overlap between automatically generated circumferences and manual ground truth. Figure 5c–f show examples for these cases.

Runtime: IC_m , SA_m and SC_m are computed on the CPU (Xeon E5-2643) and the average runtimes are 1.86 s, 0.09 s and 7.4 s respectively. S_m and FA_m are computed on the GPU and the average inference times are 1.11 s and 0.89 s.

Discussion: Because shadow areas have no solid edges and can be harder to annotated consistently than anatomy, manual segmentation can be ambiguous. Additionally, thresholding the shadow confidence map to generate a binary shadow segmentation reduces information provided by the confidence map. These two facts lead to a seemingly low DICE score when compared to current object segmentation frameworks. However, shadows are image properties rather than objects, and our final aim is to provide a confidence map, which cannot be compared quantitatively to a ground truth. The quantitative measurement in Table 1 indicates the effectiveness of the proposed method compared with the state-of-the-art method when handling complex shadow images. The qualitative results in Fig. 5 show accurate shadow detection of the proposed method and Fig. 6 demonstrate the importance of shadow detection in automatic medical image analysis.

4 Conclusion

We have presented a novel method to generate pixel-wise, shadow-focused confidence maps for 2D ultrasound. Such confidence maps can be used to identify less certain regions in images, which is important for fully automatic segmentation tasks or automatic image-based biometric measurements. We show shadow detection results of our method qualitatively and compare our method with the state-of-the-art method quantitatively. We also show the advantage of shadow confidence maps via integration into an automatic biometrics FCN. In the future we explore ways to convert our pipeline into a learn-able end-to-end approach.

Acknowledgments. Supported by the Wellcome Trust IEH Award [102431] and Nvidia Corporation.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. CoRR abs/1701.07875 (2017)
2. Baumgartner, C., et al.: SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE Trans. Med. Imaging **36**(11), 2204–2215 (2017)
3. Baumgartner, C., Koch, L., Tezcan, K., Ang, J., Konukoglu, E.: Visual feature attribution using Wasserstein GANs. CoRR abs/1711.08998 (2017)

4. Berton, F., Cheriet, F., Miron, M.-C., Laporte, C.: Segmentation of the spinous process and its acoustic shadow in vertebral ultrasound images. *Comput. Biol. Med.* **72**, 201–211 (2016)
5. Bouhemad, B., Zhang, M., Lu, Q., Rouby, J.: Clinical review: bedside lung ultrasound in critical care practice. *Crit. Care* **11**(1), 205 (2007)
6. Broersen, A., et al.: Enhanced characterization of calcified areas in intravascular ultrasound virtual histology images by quantification of the acoustic shadow: validation against computed tomography coronary angiography. *Int. J. Cardiovasc. Imaging* **32**, 543–552 (2015)
7. Coupé, P., Hellier, P., Kervrann, C., Barillot, C.: Nonlocal means-based speckle filtering for ultrasound images. *IEEE Trans. Image Process.* **18**(10), 2221–2229 (2009)
8. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.: Improved training of Wasserstein GANs. *CoRR abs/1704.00028* (2017)
9. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
10. Hellier, P., Coupé, P., Morandi, X., Collins, D.: An automatic geometrical and statistical method to detect acoustic shadows in intraoperative ultrasound brain images. *Med. Image Anal.* **14**(2), 195–204 (2010)
11. Karamalis, A., Wein, W., Klein, T., Navab, N.: Ultrasound confidence maps using random walks. *Med. Image Anal.* **16**(6), 1101–1112 (2012)
12. Klein, T., Wells, W.M.: RF ultrasound distribution-based confidence maps. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9350, pp. 595–602. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24571-3_71
13. Kremkau, F.W., Taylor, K.: Artifacts in ultrasound imaging. *J. Ultrasound Med.* **5**(4), 227–237 (1986)
14. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: *NIPS 2012*, pp. 1097–1105 (2012)
15. Lange, T., et al.: 3D ultrasound-CT registration of the liver using combined landmark-intensity information. *Int. J. Comput. Assist. Radiol. Surg.* **4**(1), 79–88 (2009)
16. NHS: Fetal anomaly screening programme: programme handbook June 2015. Public Health England (2015)
17. Rajchl, M., et al.: DeepCut: object segmentation from bounding box annotations using convolutional neural networks. *IEEE Trans. Med. Imaging* **36**(2), 674–683 (2017)
18. Sinclair, M., et al.: Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks. In: *EMBC 2018* (2018)
19. Springenberg, J., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: the all convolutional net. *CoRR abs/1412.6806* (2014)
20. Yuri, B., Vladimir, K.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1124–1137 (2004)
21. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: *CVPR 2016*, pp. 2921–2929. IEEE (2016)