林芃葦 撰

110

3

結合注意力機制與Unet3+卷積神經網路用於胎兒參數ＨＣ測量

國立臺北教育大學

資訊科學系

林芃葦

Lin-Peng Wei

指導教授：劉遠楨博士

Advisor : Liu-Yuan Zhen, Ph. D.

中華民國 110 年 3 月

2021

碩士論文

國立臺北教育大學資訊科學系

碩士論文

Department of Computer Science

College of Education

National Taipei University of Education

Master’s Thesis

結合注意力機制與Unet3+卷積神經網路

用於胎兒參數HC測量

Use Unet3+ with Attention Mechanism to calculate fetal head circumference

謝辭

能完成這篇論文，很大的一部分是我的指導教授，劉遠楨博士的幫助，指引我方向，了解了該如何自己查詢醫學論文和影像處理的方法，讓從沒接觸過醫學影像的我，能夠快速的了解，從而誕生了這篇超音波胎兒神經網路的論文。

摘要

這篇論文提出ATUnet3+類神經網路方法來計算胎兒頭圍指標(HC)，達到媲美專家的水準。

在胎兒超音波影像裡，有各種評估胎兒健康狀況的指標，本論文選擇要測量指標是Head Circumference (HC)，我們藉由類神經網路中的Unet3+並結合注意力機制來達到精準的胎兒頭部分割，得到頭部分割後，再用最小二乘法擬合橢圓，算出最後HC的輸出。

本論文方法ATUnet3+在Dice係數上的分數為97.89 ± 1.39 %、Hausdorff Distance為1.29 ± 0.83 mm、Absolution Difference為 1.90 ± 1.80 mm、Difference為 -0.51 ± 2.57 mm。

Abstract

介紹

超音波造影比起他造影方法擁有非入侵性、成本低廉、即時等優點，所以常廣泛用於婦產科當中；而缺點則是容易會有雜點、變形、扭曲，這是因為超音波是藉由收到回音來成像，而聲音在傳遞過程會有慢慢衰減。如越深的器官，聲音透射量就會越小，這是由於聲音會被組織吸收或散射或反射；吸收散射造成影像sign-to-noise低；反射迴盪回音造成影像變形；相同聲阻抗的不同組織造成分界難以辨別；這是超音波先天上的特性。

而除了上述的困難點，還有一個也很棘手的問題就是測量結果會依據超音波師們的判斷經驗不同及手法操作不同，所得到的造影結果也會有大大的不同，然而訓練經驗有素的超音波師是需要成本時間的，比較偏鄉或是發展中的國家普遍缺乏，很多母親都是在沒有超音波監控下生產，有99%生產死亡都是發生在發展中國家 [1] ，所以能夠自動化測量超音波胎兒的研究是非常重要的議題。

在胎兒超音波造影中有各種胎兒生物參數可以去判斷胎兒生長狀況及異常，例如：biparietal diameter (BPD)、head circumference (HC)、abdominal circumference (AC) 和femur length (FL) 等等。本論文所選擇評估的胎兒參數是HC，HC可以用來計算胎齡和生長有無異常。

過去已經有使用傳統的機器學習方法來自動化偵測胎兒HC，但都有些缺點，如Carneiro等人[2] 使用probabilistic boosting tree去預估和分類超音波分割結果，雖然結果不錯，不過缺點是要訓練樹需要有大量專家標記完好且有點複雜的資料；Li等人 [3] 使用隨機森林去定位胎兒頭部位置然後使用phase symmetry 和去擬合橢圓計算HC，但是這方法需要先知道胎齡及超音波掃描的深度；Lu等人 [4] 採用 K-mean 和 morphological operation 來取得頭部片段，然後用 iterative randomized Hough transform 去擬合頭部橢圓，但問題是K-mean不能有效的抽出頭部片段，這是由於超音波的雜訊太多，讓後續的 Hough transform 不穩定；Stebbing 和 McManigle [5] 提出使用 boundary fragment model 方法再丟給隨機森林做邊緣偵測，然而這個方法會因為超音波造影到其他解剖構造如胎盤的影像強度和胎兒頭部相似，影響此方法的穩定性。

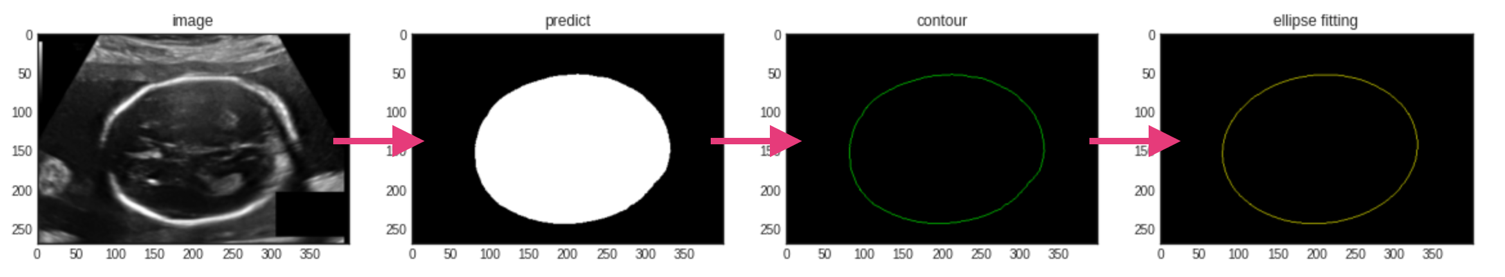
近幾年隨著神經網路的崛起，人們開始把神經網路用臨床研究上，在之中CNN對於圖片有加良好的分割結果，所以也被用到醫學影像上，如Matthew Sinclair等人 [6] 提出使用全卷積網路(FCNs)來測量胎兒HC，Zahra Sobhaninia 等人 [7] 提出使用多尺度的Link-Net來測量胎兒HC，結果顯示在各指標上都有不錯的分數；目前這些網路的架構都是採用類似Unet [8] 的架構的思想，Unet的特色是編碼在解碼，並用跳躍連接來讓解碼層考慮到因為降採樣而消失的細部特徵。

在Unet之後，Zongwei Zhou等人 [9] 改良了跳躍連結的部分，提出Unet++，改變了Unet跳躍連接的方式，採用密集的跳躍連接，達成類似集成的方式，大大提升了UNet的各項分割指標，但是由於密集的跳躍連接使得需訓練的參數大幅增加，增加了訓練成本，此外，每個解碼層節點並沒有考量到各層編碼層節點的輸出，沒有考量到多尺度的特徵圖，無法去融合淺層特徵考慮到細部特徵，因此Huimin Huang等人 [10] 提出了Unet3+。

Unet3+也改變了跳躍連接的部分，讓每一個編碼層的特徵圖，都會併接到解碼層，達到多尺度的考量，並且由於固定通道數，需訓練的參數量也比Unet++和Unet少，在分割結果上則比Unet++稍稍提升。

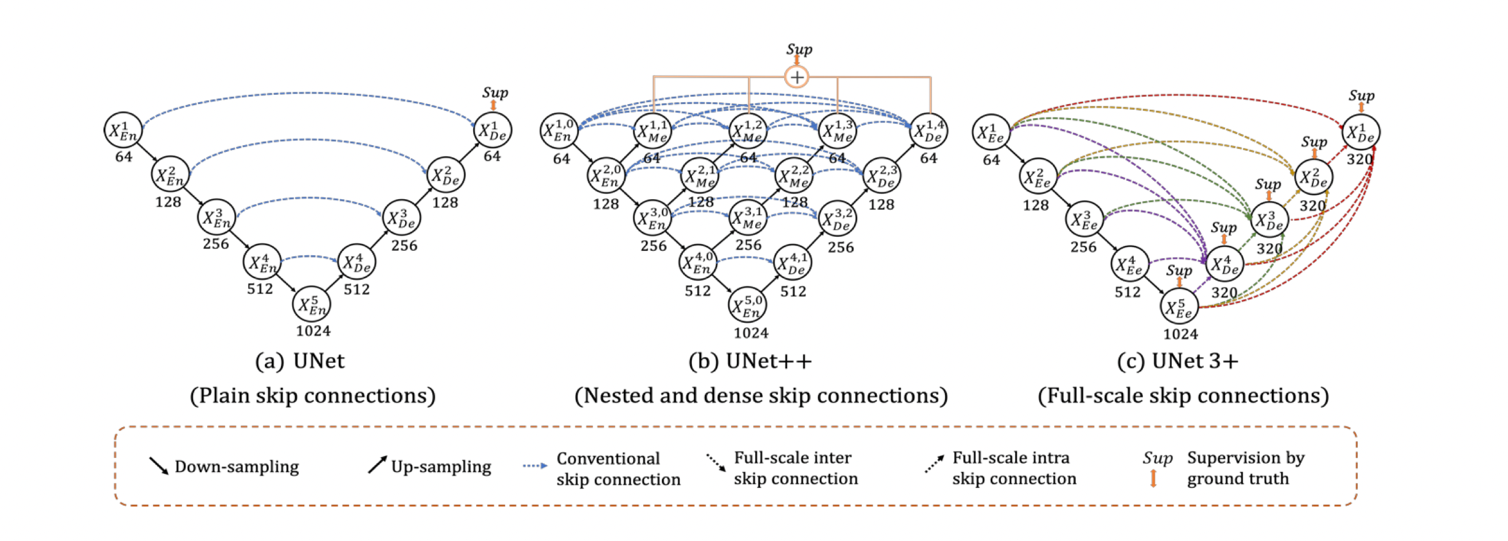
本論文所提出的方法是基於Unet3+上引入注意力機制成為ATUnet3+，然後使用混合loss funtion，結合BCE + MS-SSIM + DICE三種，

得到精確的分割結果後，再使用最小二乘橢圓擬合計算出HC，最後使用Dice係數、Hausdorff Distance、Absolution Difference、Difference做評估的指標。

方法

流程圖如Fig.1 透過本論文提出的ATUnet3+取得胎兒頭部ROI，然後抽取邊緣再用橢圓擬合得到最終輪廓。

Unet3+介紹



Unet3+由Huimin Huang等人 [10] 提出，將編碼層各個尺度的特徵圖，跳躍併接到每個解碼層，讓解碼層節點可以考慮多尺度的特徵，輸出較好的分割結果，而且訓練參數少於Unet和Unet++，可以達到更快收斂及運行速度更快的效果。

Unet 解碼層所需訓練參數計算：

K：卷積核大小，本論文用的數值是3。

d： 的channels數。

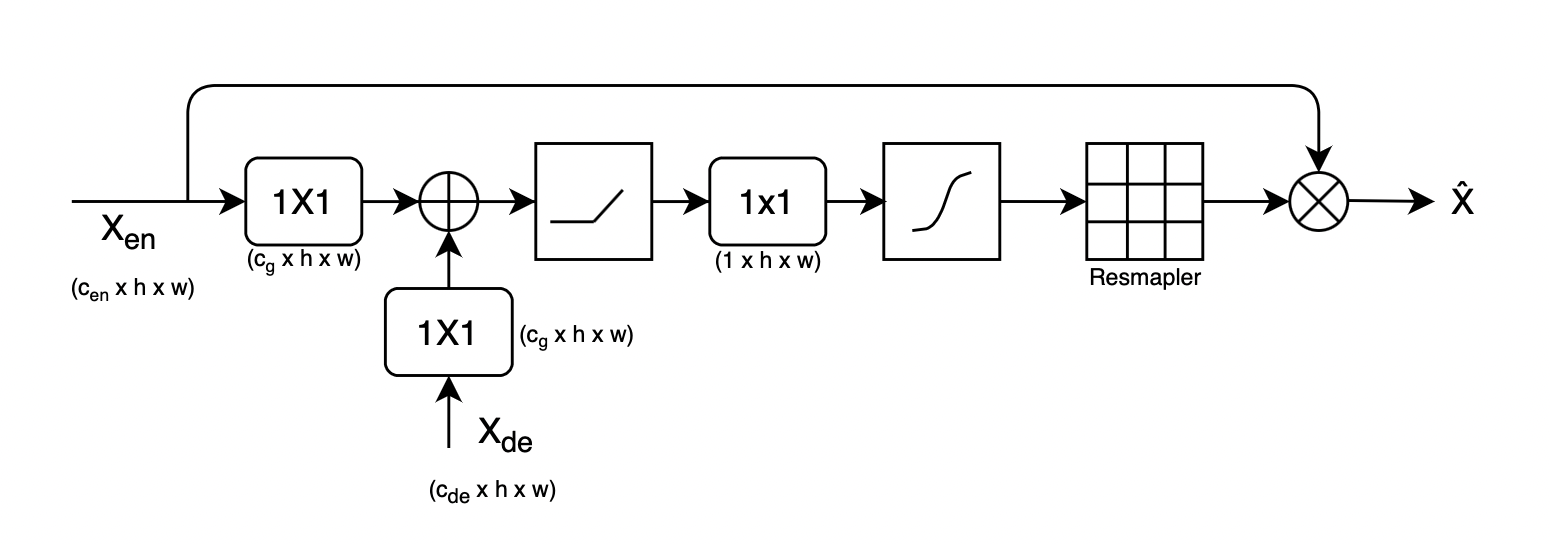
Unet++ 解碼層所需訓練參數計算：

：是中間層節點。

Unet3+ 解碼層所需訓練參數計算：

由於他們Encoder的結構一樣，所以需要的訓練參數量是一樣的，主要在於Unet++的中間層和Decoder的差別，所以我們比較他們最後連接到Decoder所需要的訓練參數就好；Unet和Unet3+差別主要就是channel數，透過以上公式計算每一層解碼層的節點相加後，可以發現Unet3+比Unet所需訓練參數更低，這是由於Unet3+的解碼層節點得channel數都是固定320，而Unet和Unet++都是 channels數。

注意力機制介紹



注意力機制常用於自然語言處理、知識圖譜、影像分析，如image caption [11]、機器翻譯[12]、分類任務[13][14]等等，而在醫學影像分割上，Ozan Oktay等人在Unet上應用了注意力機制[15]，該注意力機制用於跳躍連接中，讓機器要併接的特徵圖中給重要區域比較大的權重，抑制不相關的區域，由於深層的節點能獲得更多高級抽象特徵訊息，所以用深層的節點來監督上一層編碼層節點的輸出。注意力公式為以下，輸出加權後的特徵圖：

：編碼層輸出。

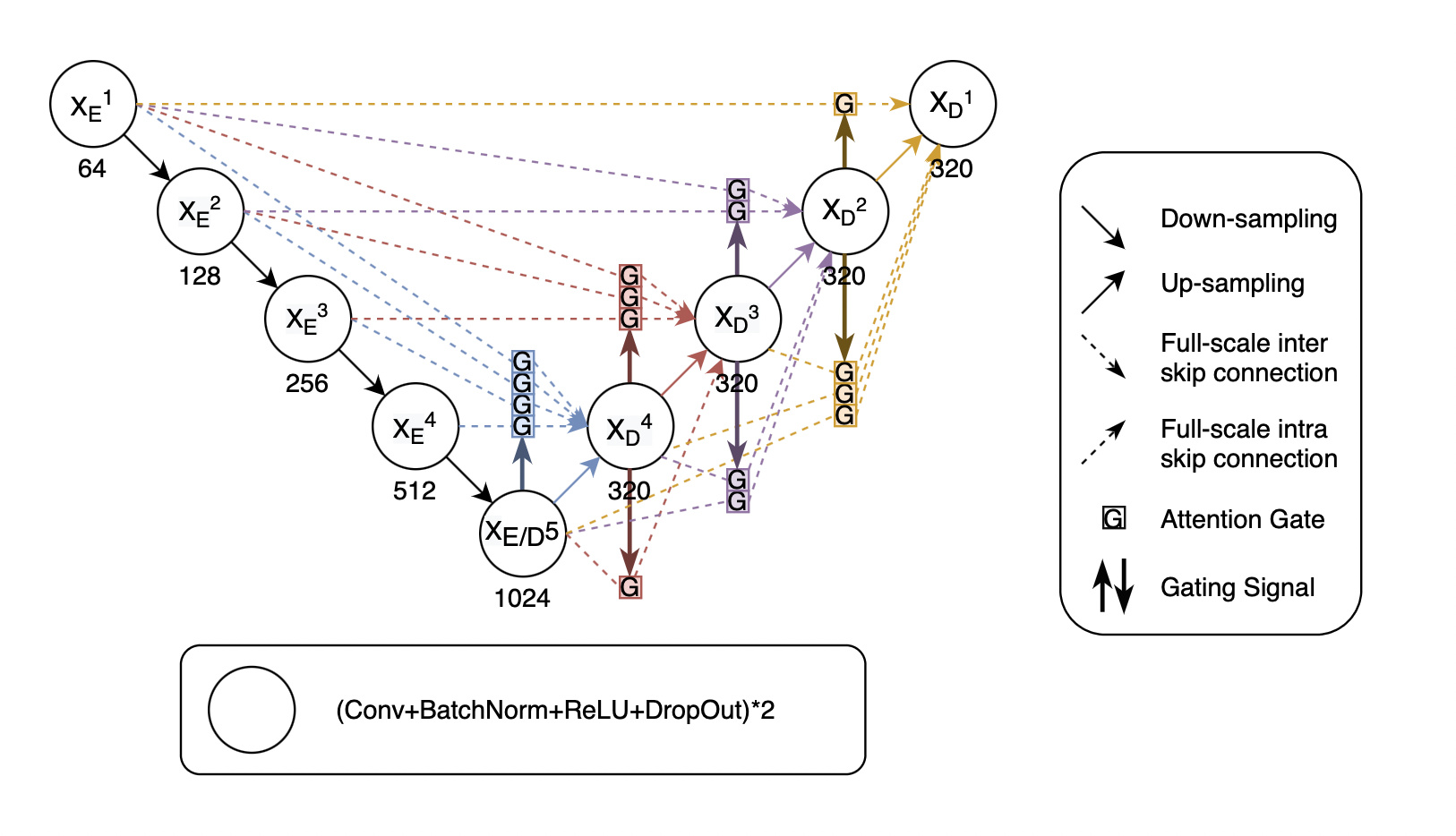
：解碼層輸出。

：1x1卷積，用來讓和的channels數目一樣，才可以做矩陣相加。

：Relu激活函式

：Sigmoid激活函式

本論文提出的ATUnet3+介紹



通過結合以上兩點，我們提出了ATUnet3+(Attetion Unet3+)的架構Fig.3，改善多尺度特徵圖的跳躍連接直接到解碼層，讓機器去學習這些特徵圖哪些部份是要著重的，從而

達到不錯的分割結果。

我們對每一個編碼層的跳躍連結的輸出使用注意力機制來讓機器學習哪些區域是重要的，再並接到解碼層上，ATUnet3+解碼層的節點輸出的公式由以下得出：

：是第i層的解碼層輸出。

：是總共有多少個解碼層。

是對特徵圖做併接。

是通過注意力機制後輸出。

是做卷積操作。

是做下採樣動作，這裡使用MaxPool。

是做上採樣動作，這裡使用反捲機和雙線性來上採樣。

Loss function介紹

我們這裡採用混合loss funtion 來評估，結合BCE + MS-SSIM + DICE三種Loss function

分別代表，像素層級、片段層級、重疊層級，各方面來幫助我們訓練。

MS-SSIM(Multi Scale Structural Similarity index measure)介紹：

SSIM [16] 是對兩張圖的相似度做評估的指標，綜合了亮度、對比度、結構方面，但是由於圖像特徵的空間域在統計上是非平穩過程，以及人類再看一張圖時，我們都會看一小部分的區域，才能看得清晰，所以實際上會使用MSSIM (Mean SSIM)。

MSSIM是用個滑動視窗NxN大小，計算區域的SSIM，對每個區域都滑過後，取平均得到，

MS-SSIM [17] 就是多尺度版本的MSSIM，會將要比較的圖像縮半，並對每個尺度都有個權重，在此根據論文我們設權重為[0.0448, 0.2856, 0.3001, 0.2363, 0.1333]，用5個尺度。

MS-SSIM的範圍為[0,1]，值越接近1越好，而這裡我們要算的是Loss function，所以。

SSIM亮度(luminance) 評估公式：

SSIM 對比度(contrast) 評估公式：

SSIM 結構 (contrast) 評估公式：

：x的平均值。

：y的平均值。

：x的標準差。

：y的標準差。

：共變異數。

：值為0.01

：值為0.03

：值為255，是像素8bit值的最大範圍。

：值為。

：值為。

：值為。

綜合以上得到SSIM公式：

：可以自己設定每個函式的權重，這裡都設為1

MSSIM是取SSIM每個滑動視窗的平均，利用NxN大小的滑動視窗滑完整張圖，得到總數為M個視窗，公式為：

MS-SSIM透過不斷(這裡論文縮小4次，5個尺度)縮小一半比較的圖片，可以得到不同尺度的相似程度，其公式為：

值為1，代表沒有縮小，通常對亮度相似度計算只會計算一次。

通常令

論文給出了5個尺度適合的權重，分別是

BCE(Binary Cross Entropy)介紹：

BCE是CrossEntropy [18] 的一個特例，用來評估兩個機率分佈的差異，常用來當分類或是分割任務的loss function。這裡我們用來將預測的圖片的像素點機率和超音波師的標註的像素做評估，而因為我們只有兩種分類，不是頭部就是背景，所以採用BCE，他的公式是：

n：為預測圖像的所有像素點。

x：我們預測的圖像的每個像素是胎兒頭部的機率。

y：為超音波師標注的是胎兒頭部的機率，不是1就是0。

Dice係數(Sørensen–Dice coefficient)介紹：

Dice 係數能用來評估兩張圖片的相似程度，本論文用來評估我們預測的圖像與超音波師的標注的重疊率，公式是：

x：我們預測的圖像的每個像素是胎兒頭部的機率。

y：為超音波師標注的是胎兒頭部的機率，不是1就是0。

Dice係數範圍為[0,1]之間，越接近1代表我們預測的分割圖和超音波師的標註重疊率越高，而因為這裡loss是要最小化，所以

結合上面3種loss function得到我們的混合loss function：

實驗與結果

參考資料

[1] World Health Organization, UNICEF, UNFPA, The World Bank, the United Nations Population Division. Trends in maternal mortality: 1990 to 2013: estimates by WHO, UNICEF, UNFPA, The World Bank and the United Nations Population Division. 2014.

[2] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, “Detection and measurement of fetal anatomies from ultrasound images using a constrined probabilistric boosting tree,” IEEE Transactions on Medical Imaging, vol. 27, pp. 1342-1355, 2008.

[3] J. Li, Y. Wang, B. Lei, J.-Z. Cheng, J. Qin, T. Wang, S. Li, and D. Ni, “Automatic Fetal Head Circumference Measurement in Ultrasound using Random Forest and Fast Ellipse Fitting,” IEEE journal of Biomedical and Health Information, DOI:10.1109/JBHI.2017.2703890, 2017.

[4] W. Lu, J. Tan, and R. Floyd, “Automated fetal head detection and measurement in ultrasound images by iterative randomized hough transform,” Ultrasound in medicine & biology, vol. 31, no. 7, pp. 929–936, 2005.

[5] R. V. Stebbing and J. E. McManigle, “A boundary fragment model for head segmentation in fetal ultrasound,” Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI 2012, pp. 9–11, 2012.

[6] Matthew Sinclair, Christian F. Baumgartner, Jacqueline Matthew, Wenjia Bai, Juan Cerrolaza Martinez, Yuanwei Li, Sandra Smith, Caroline L. Knight, Bernhard Kainz, Jo Hajnal, Andrew P. King, Daniel Rueckert, “Human-level Performance On Automatic Head Biometrics In Fetal Ultrasound Using Fully Convolutional Neural Networks”

[7] Zahra Sobhaninia, Shima Rafiei, Ali Emami, Nader Karimi, Kayvan Najarian, Shadrokh Samavi, S.M.Reza Soroushmehr, “Fetal Ultrasound Image Segmentation for Measuring Biometric Parameters Using Multi-Task Deep Learning”

[8] Olaf Ronneberger, Philipp Fischer, Thomas Brox,“U-Net: Convolutional Networks for Biomedical Image Segmentation”

[9] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation”

[10] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, Jian Wu, “UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation”

[11] Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, Lei Zhang, “Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering”

[12] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate”

[13] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, Yoshua Bengio, “Graph Attention Networks”

[14] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, Xiaoou Tang, “Residual Attention Network for Image Classification”

[15] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert, “Attention U-Net: Learning Where to Look for the Pancreas”

[16] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, IEEE, and Eero P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity”

[17] Zhou Wang1, Eero P. Simoncelli1 and Alan C. Bovik2, “MULTI-SCALE STRUCTURAL SIMILARITY FOR IMAGE QUALITY ASSESSMENT”

[18] Ma Yi-de, Liu Qing, Qian Zhi-bai,“Automated image segmentation using improved PCNN model based on cross-entropy”