Neonatal Face Segmentation with and without Clinical Devices using SAM

Pedro Henrique Silva Domingues*, Tatiany Marcondes Heiderich*†,
Marina Carvalho de Moraes Barros†, Ruth Guinsburg† and Carlos Eduardo Thomaz*
* FEI, São Paulo, Brazil
† UNIFESP, São Paulo, Brazil

***Abstract*—Facial expression analysis has been widely used as one of the main approaches for pain diagnosis, both by humans and computing systems. However, in clinical practice, newborns who remain hospitalized in Neonatal Intensive Care Units often have devices connected to their faces, such as enteral/gastric probes, orotracheal intubation tubes, and phototherapy goggles, which hinder the visualization of facial regions making the proper diagnosis of pain much harder in practice. Therefore, to address this issue, we have evaluated the state-of-the-art Segment Anything Model (SAM) tool combined with RetinaFace for segmentation of 2D face images of neonates, including free faces and the ones with devices connected, against a simple and traditional landmark method, and a recently proposed deep neural network fine-tuned for face segmentation under occlusion. SAM performed comparatively better than the other two models for both no occlusion and high occlusion 2D face images, scoring on average impressively 0.98 and 0.91 at the standard dice similarity coefficient respectively.**

I. Introduction

Numerous studies have been conducted to understand the approach of healthcare professionals in assessing neonatal pain [1]. These investigations range from describing the perceptions of these professionals [2], [3] to exploring the visual tracking employed by them [4], [5], [6], [7]. In recent years, in addition to conventional methods [8] of pain assessment in this population, computational methods [9] have been developed with the aim of automating this process and assisting professionals in decision-making.

Facial expression analysis has been widely used as one of the primary approaches for pain diagnosis, both by human evaluators [10], [11] and computer systems [12], [13], [14]. However, some assessment scales employed by healthcare professionals allow for the identification of specific facial characteristics [10] that discriminate the presence or absence of pain, whereas recent computational methods only perform a global analysis of the face [12], [13], [14]. In other words, unlike the clinical scales used by health professionals that allow identifying specific facial regions that help to infer the presence or absence of pain [10], the current computational methods perform only a holistic or global analysis of the neonatal face. Therefore, such methods do not address the practical difficulties of identifying pain in neonates who remain with devices attached to their faces [8], including the enteral/gastric tube attachment, orotracheal intubation attachment, and phototherapy goggles.

In this work, we propose and implement a computational method for neonatal face segmentation [15], based on the state-of-the-art Segment Anything Model (SAM) [16] combined with the well-known RetinaFace [17] tool, as an initial step to address this issue. For evaluation, we compare its performance against a simple segmentation approach based on landmark localization [18] and a recently proposed deep neural network fine-tuned for face segmentation under occlusion, named DeepLabV3+ [19], using two separate datasets of neonatal face images with and without clinical devices.

II. Materials

We have used the UNIFESP face database [20], which contains 122 and 238 images of neonates before and after a painful procedure, respectively, with image resolution of 450x233 and no facial occlusions. All images are labeled accordingly to the Neonatal Facial Coding System (NFCS) [10]. In addition to this dataset, we have created another set of neonatal face images, named here as the occlusion dataset, using only 10 images of 5 different infants (two of each) with medical equipments that partially obstruct their faces. All these 10 images have image resolution of 2322x4128 and, as well as UNIFESP dataset, were captured with parent consent and after approval of the Research Ethics Committee (1.150.901, 07/15/2015).

III. Experimental Methodology and Results

Firstly, we used SAM as a tool to help generate masks for all the 360 UNIFESP and 10 occlusion 2D face images. All masks were then manually adjusted to be used as ground truth. To infer the bounding box coordinates we applied the RetinaFace model.

All the UNIFESP images were successfully processed, but out of the 10 images with occlusion, we were able to extract the correct coordinates for only 8. One image was still usable, but the bounding box covered part of the infant's chest as well, and one failed to process since no coordinates were found due to a large amount of occlusion. Using the bounding boxes as input for SAM, we generated masks for each image and, since the model always returns three masks per image in order to resolve ambiguity, with one internal score each, we have selected the resulting mask with the highest internal score automatically.

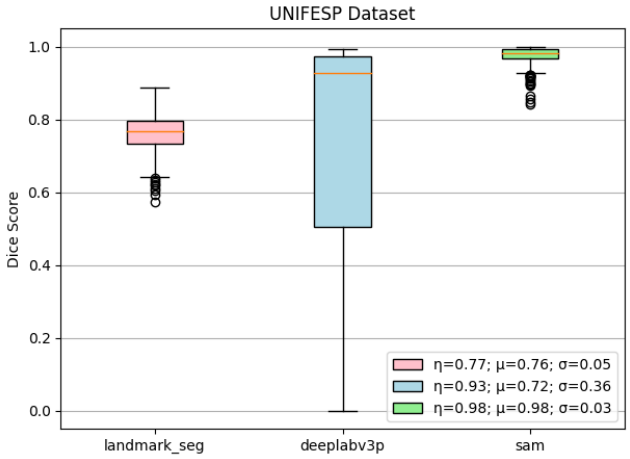


Fig. 1. Dice scores for all methods in the UNIFESP dataset.

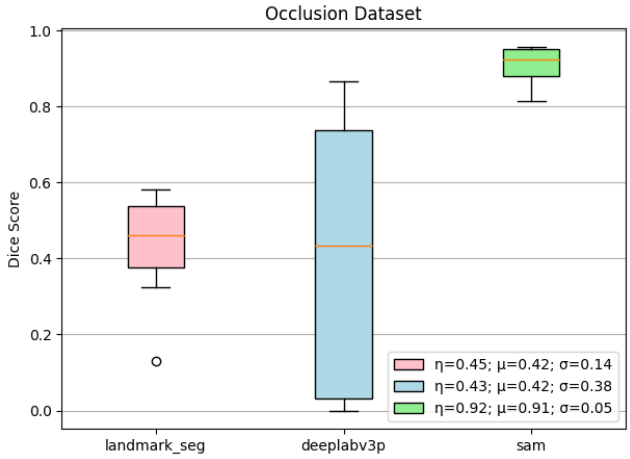


Fig. 2. Dice scores for all methods in the occlusion dataset.

For comparison, two other face segmentation methods were applied in the same datasets.

1) **Landmark segmentation**: Consists in the localization of 106 facial landmarks [18], followed by the creation of a convex hull, subsequently filled to generate the facial mask. The method is constrained by two primary limitations, namely the absence of landmarks for the forehead and the inherent inability to discern obstructive elements. However, despite these limitations, it remains a solid base for comparative analysis, presenting high performance on images devoid of obstructions;
2) **DeepLabV3+**: Fourth iteration of the DeepLab model (a convolution based neural network) developed for semantic segmentation by the google research group [21]. The model used in this paper is a version fine-tuned for face segmentation in the presence of obstruction [19]. It is important to point out that the dataset used by [19] for the fine-tune process consisted mostly of images of adults.

To quantify the effectiveness of the segmentation carried out, the standard dice similarity coefficient (score) was calculated for all masks generated, as presented in Figures 1 and 2.

A. Landmark segmentation

As shown in Figures 1 and 2 the landmark segmentation method has a high score for the UNIFESP dataset, achieving a mean of 0.76 for the dice similarity coefficient, which could be higher if not for the absence of forehead landmarks, as shown in Figure 3c and 4c. When used in the occlusion dataset, a noticeable lowering in the score can be seen, which, despite the landmark problem, occurs by wrong estimation of landmark position, resulting in distorted or rotated masks that don't perfectly align with the face.

B. DeepLabV3+ fine-tuned

Observing the DeepLabV3+ fine-tuned results in Figure 1, it is clear that, for the UNIFESP dataset a higher mean than the landmark segmentation is presented but a standard deviation more than seven times higher makes it a much more unstable option. The same problem can be seen for the occlusion dataset, Figure 2, and, after manually evaluating the generated masks, we concluded that in most cases the masks segmented part of the face together with parts of the body (Figure 4d), but only on situations in which the skin is visible. Five out of ten images from the occlusion dataset presented masked most of the image correctly scoring values higher than 0.7 and the remaining presented a low masked pixel count, not segmenting the face and resulting in scores lower than 0.25, which elevated the standard deviation to 0.38 for this dataset.

C. SAM

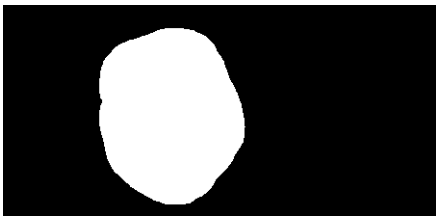
With an average score of 0.98 and a standard deviation of 0.03 for the UNIFESP dataset, SAM was the best performer from the three methods in the unobstructed face scenario. Furthermore, droping the mean to 0.91 and increasing the standard deviation by only 0.02 when applied to the occlusion dataset, this method can be considered robust to the facial obstruction by medical devices, even though it was not fine-tuned for this type of image.

After a qualitative analysis of the masks generated by SAM, the problems consisted in missing chunks of small areas in the mask, as noticeable in Figure 5d, and parts of the neck and chest segmented in conjunction with the face, as shown in Figures 3e. Some cases presented the segmentation of the medical devices instead of the face as one of the three output masks, as illustrated in Figure 5, which may prove to be a problem when applied to a larger dataset.

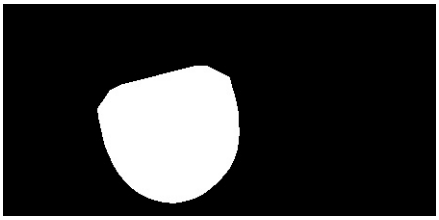
IV. Conclusion

In this paper, we proposed and implemented a comparative analysis between SAM and two other segmentation models, a traditional landmark one and the DeepLabV3+ recently developed, with the aim of experimentally investigating the most promising methodology for neonatal face segmentation and further pain assessment classification. All models were

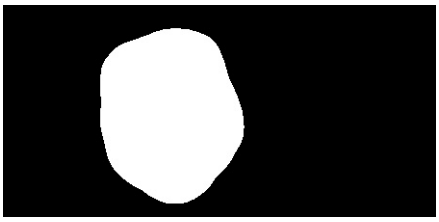
(a) Original image



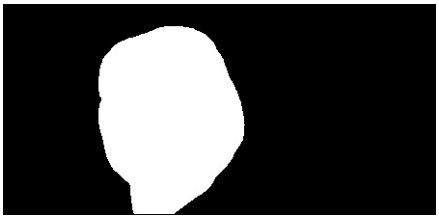
(b) Ground truth mask



(c) Landmark segmentation



(d) DeepLabV3+



(e) SAM highest score mask

Fig. 3. Segmentation of an image from the UNIFESP dataset.

evaluated against both a free face dataset and a high facial obstruction dataset. SAM performed impressively well for both no occlusion and high occlusion 2D face images, scoring on average 0.98 and 0.91 at the standard dice similarity coefficient, with a standard deviation lower than and equal to 0.05 respectively. The main SAM's drawback, however, was to separate exclusively the face segmentation from the neck and chest ones. For future work, we will focus on fine-tuning SAM to address this drawback and evaluate complementary computational methods to turn this binary segmentation into a semantic map of regions of interest like NFCS using larger datasets.

ACKNOWLEDGMENT

The authors would like to thank the financial support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - Finance Code 001), Fundação Educacional Inaciana Padre Sabóia de Medeiros (FEI), and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - 2018/13076-9).

REFERENCES

[1] M. García-Rodríguez, S. Bujan-Bravo, R. Seijo-Bestilleiro, and C. Gonzalez, "Pain assessment and management in the newborn: A systematized review," *World journal of clinical cases*, vol. 9, pp. 5921–5931, 07 2021.

[2] T. M. Heiderich, M. C. d. M. Barros, and R. Guinsburg, "Concordância interavaliadores na identificação de faces de dor de recém-nascidos a termo e pré-termo tardio: estudo transversal," *BrJP*, vol. 3, pp. 348–353, 2020.

[3] G. De Clifford Faugère, M. Aita, N. Feeley, S. Colson *et al.*, "Nurses' perception of preterm infants' pain and the factors of their pain assessment and management," *The Journal of Perinatal & Neonatal Nursing*, vol. 36, no. 3, pp. 312–326, 2022.

[4] G. V. T. d. Silva, M. C. d. M. Barros, J. d. C. A. Soares, L. P. Carlini, T. M. Heiderich, R. N. Orsi, R. d. C. X. Balda, C. E. Thomaz, and R. Guinsburg, "What facial features does the pediatrician look to decide that a newborn is feeling pain?" *American Journal of Perinatology*, vol. 40, no. 08, pp. 851–857, 2021.

[5] J. d. C. A. Soares, M. C. d. M. Barros, G. V. T. d. Silva, L. P. Carlini, T. M. Heiderich, R. N. Orsi, R. d. C. X. Balda, P. A. S. O. Silva, C. E. Thomaz, and R. Guinsburg, "Looking at neonatal facial features of pain: do health and non-health professionals differ?" *Jornal de Pediatria*, vol. 98, pp. 406–412, 2022.

[6] M. C. d. M. Barros, C. E. Thomaz, G. V. T. da Silva, J. do Carmo Azevedo Soares, L. P. Carlini, T. M. Heiderich, R. N. Orsi, R. d. C. X. Balda, P. A. S. O. Silva, A. Sanudo *et al.*, "Identification of pain in neonates: the adults' visual perception of neonatal facial features," *Journal of Perinatology*, vol. 41, no. 9, pp. 2304–2308, 2021.

[7] R. Orsi, L. Carlini, T. Heiderich, G. Silva, J. Soares, R. Balda, M. Barros, R. Guinsburg, and C. Thomaz, "Visual attention during neonatal pain assessment: A 2-second exposure to a facial expression is sufficient," *Authorea Preprints*, 2022.

[8] A. Llerena, K. Tran, D. Choudhary, J. Hausmann, D. Goldgof, Y. Sun, and S. Prescott, "Neonatal pain assessment: Do we have the right tools?" *Frontiers in Pediatrics*, vol. 10, 02 2023.

[9] S. Gkikas and M. Tsiknakis, "Automatic assessment of pain based on deep learning methods: A systematic review," *Computer Methods and Programs in Biomedicine*, vol. 231, p. 107365, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260723000329

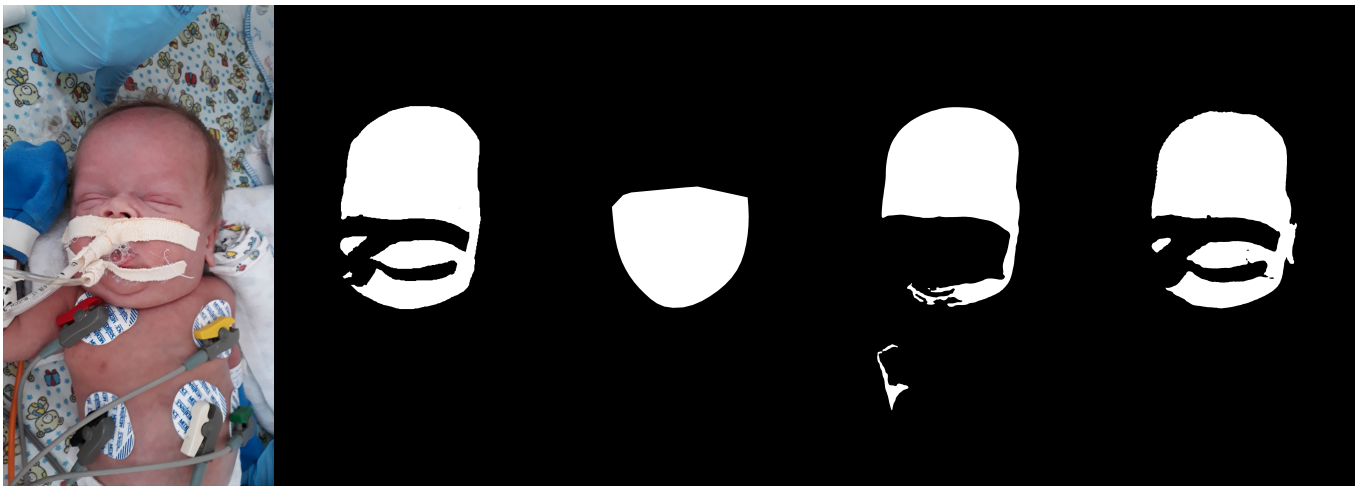
[10] R. V. Grunau and K. D. Craig, "Pain expression in neonates: facial action and cry," *Pain*, vol. 28, no. 3, pp. 395–410, 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/030439598790073X

[11] B. J. Stevens, S. Gibbins, J. Yamada, K. Dionne, G. Lee, C. Johnston, and A. Taddio, "The premature infant pain profile-revised (pipp-r): initial validation and feasibility," *The Clinical journal of pain*, vol. 30, no. 3, pp. 238–243, 2014.

[12] G. Zamzmi, R. Paul, M. S. Salekin, D. Goldgof, R. Kasturi, T. Ho, and Y. Sun, "Convolutional neural networks for neonatal pain assessment," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 3, pp. 192–200, 2019.

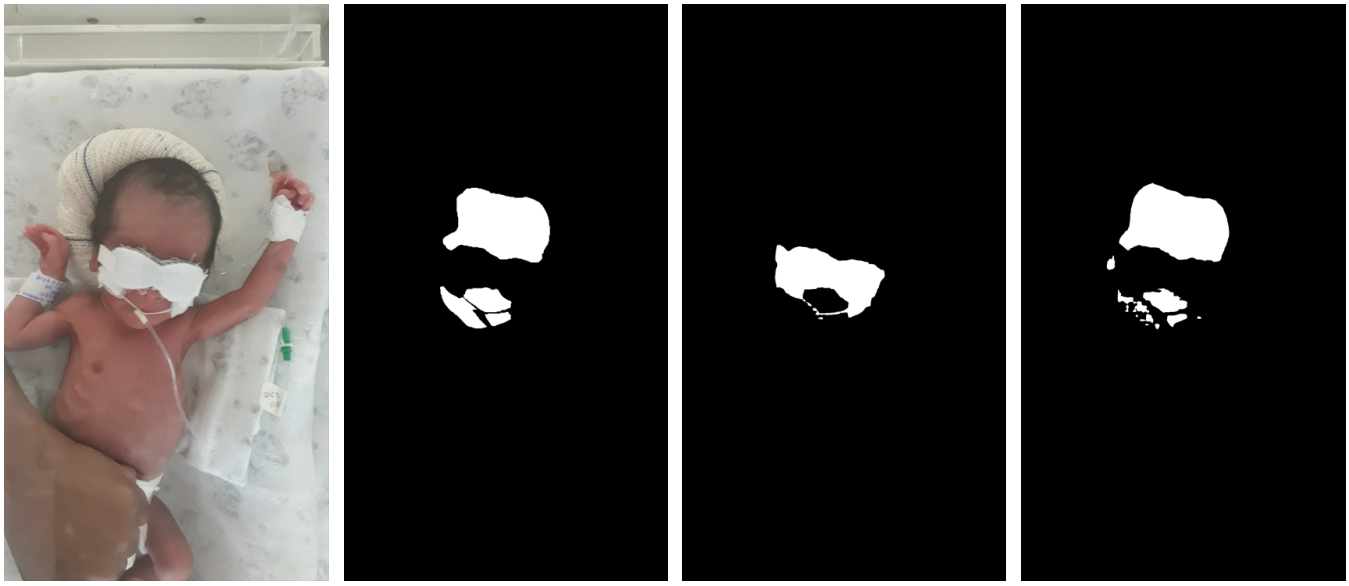
[13] L. P. Carlini, L. A. Ferreira, G. A. Coutrin, V. V. Varoto, T. M. Heiderich, R. C. Balda, M. C. Barros, R. Guinsburg, and C. E. Thomaz, "A convolutional neural network-based mobile application to bedside neonatal pain assessment," in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2021, pp. 394–401.

[14] P. H. S. Domingues, R. M. M. da Silva, I. J. Orra, M. E. Cruz, T. M. Heiderich, and C. E. Thomaz, "Neonatal face mosaic: An areas-of-interest segmentation method based on 2d face images," in *Anais do XVII Workshop de Visão Computacional*. SBC, 2021, pp. 201–205.



(a) Original image (b) Ground truth mask (c) Landmark segmentation (d) DeepLabV3+ (e) SAM highest score mask

Fig. 4. Segmentation of an image from the occlusion dataset using all methods.



(a) Original image (b) Ground truth mask (c) SAM phototherapy goggles segmentation (d) SAM face segmentation

Fig. 5. Segmentation of both both face and googles by SAM from the same input.

[15] Y. S. Dosso, D. Kyrollos, K. J. Greenwood, J. Harrold, and J. R. Green, "Nicuface: Robust neonatal face detection in complex nicu scenes," *IEEE Access*, vol. 10, pp. 62 893–62 909, 2022.

[16] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.

[17] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5202–5211.

[18] Y. Liu, H. Shen, Y. Si, X. Wang, X. Zhu, H. Shi, Z. Hong, H. Guo, Z. Guo, Y. Chen, B. Li, T. Xi, J. Yu, H. Xie, G. Xie, M. Li, Q. Lu, Z. Wang, S. Lai, Z. Chai, and X. Wei, "Grand challenge of 106-point facial landmark localization," *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 613–616, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:148571603

[19] K. T. R. Voo, L. Jiang, and C. C. Loy, "Delving into high-quality synthetic face occlusion segmentation datasets," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[20] T. Marcondes Heiderich and A. Leslie, "Neonatal procedural pain can be assessed by computer software that has good sensitivity and specificity to detect facial movements," *Acta Paediatrica*, vol. 104, 11 2014.

[21] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 833–851.