Pain Evaluation in Video using Extended Multitask Learning from Multidimensional Measurements

Xiaojing Xu
Department of Electrical and Computer Engineering
UC San Diego, CA, USA

Virginia R. de Sa
Department of Cognitive Science and Halıcıoğlu Data Science Institute
UC San Diego, CA, USA

Previous work on automated pain detection from facial expressions has primarily focused on frame-level objective pain metrics, such as the Prkachin and Solomon Pain Intensity (PSPI). However, the current gold standard pain metric is the visual analog scale (VAS), which is self-reported at the video level.

This abstract is a compressed version of [1]. In this work, we propose and analyze a multitask multidimensional-pain model to directly evaluate VAS in video. Our model consists of three stages: (1) a VGGFace neural network multitask learning model to predict frame-lavel PSPI and facial action units; (2) a fully connected neural network to estimate sequence-level pain scores from these frame-level predictions, where again we use multitask learning to learn multidimensional pain scales instead of VAS alone; and (3) an optimal linear combination of the multidimensional sequence-level pain predictions to obtain a final estimation of VAS. The structure of our model is shown in Figure 1.

We show on the UNBC-McMaster Shoulder Pain dataset that our multitask multidimensional-pain method achieves state-of-the-art performance with an MAE of 1.95 and ICC of 0.43. When combined with human observer estimations, our model improves their MAE from 1.76 to 1.58. Trained on the UNBC-McMaster dataset and applied directly with no further training or fine-tuning on a separate dataset of facial videos recorded during post-appendectomy physical exams, our model also outperforms previous work by 6% on the Area under the ROC curve metric (AUC).

We thank NVIDIA Corporation for the donation of the Titan V GPU used for this research

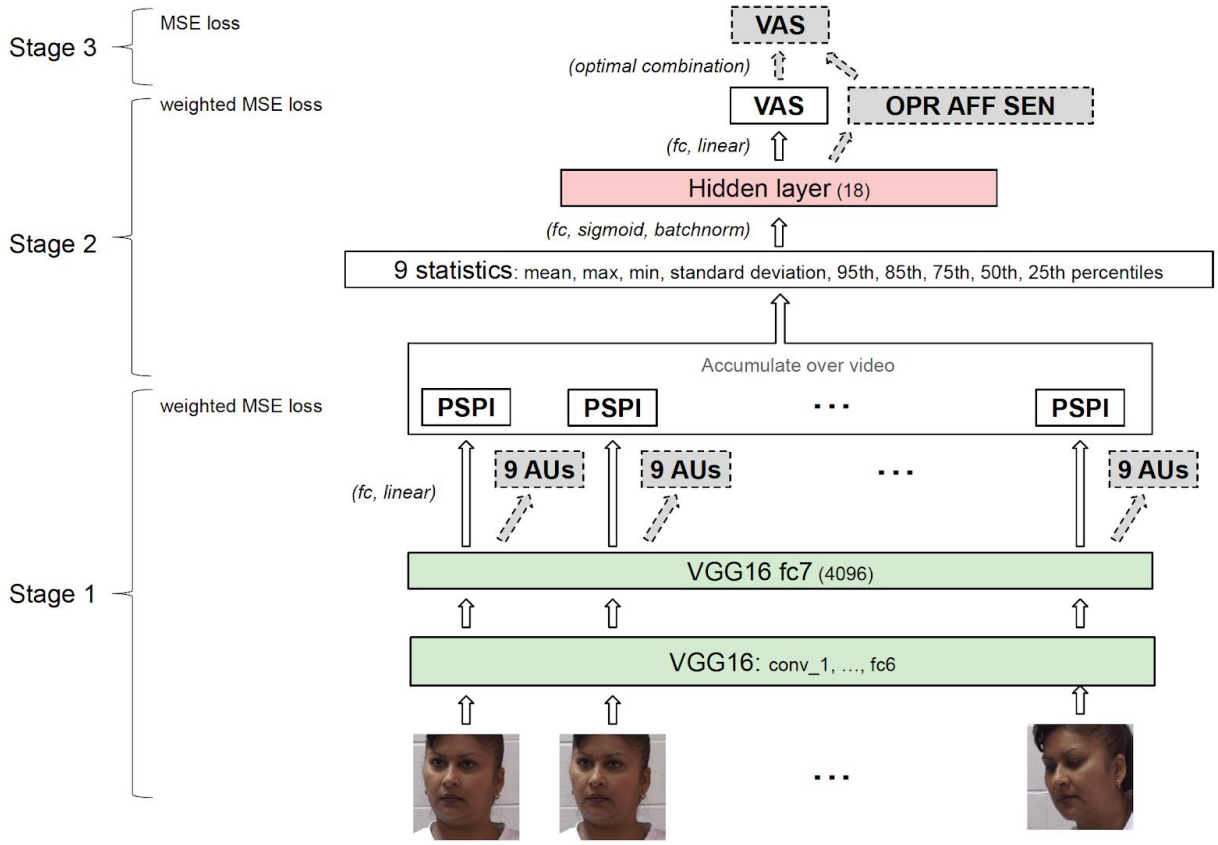

Figure 1. The proposed three-stage structure. The baseline model is represented by solid blocks, and shaded blocks with dashed outlines show added parts in multitask learning and ensemble learning with multidimensional pain scales. During training, Stage 1 is first trained with batches of frames and used to predict a sequence of PSPI scores. Then Stage 2 is learnt using batches of video features obtained from PSPI sequences. The network can't be trained end-to-end due to limited GPU memory.

[1] Xu, Xiaojing, Jeannie S. Huang, and Virginia R. De Sa. "Pain Evaluation in Video using Extended Multitask Learning from Multidimensional Measurements." In Machine Learning for Health ML4H at NeurIPS 2019, pp. 141-154. 2020.