MOONEY FACE CLASSIFICATION AND PREDICTION BY LEARNING ACROSS TONE

Tsung-Wei Ke*†    Stella X. Yu*†    David Whitney*

* UC Berkeley / †ICSI

ABSTRACT
Mooney faces are special two-tone images that elicit a rich impression of identity and facial expression in human observers. While Mooney faces are important, there exist only a small number of instances hand-crafted from source photos which are often no longer available.

We first apply deep learning methods to generate a plausible Mooney face automatically from any face photo. We are then able to create a large-scale face dataset with paired grayscale and two-tone images.

We then study how well two-tone versions make face predictions, using conditional Generative Adversarial Networks. We show that faces predicted from Mooney images bear striking resemblance to source photos, and they are better than two-tone images obtained by global intensity thresholding. We also demonstrate remarkable face predictions from very low resolution surveillance photos. Our findings reveal great potentials of combining deep learning and Mooney faces for more effective face recognition in a wide range of conditions.

Index Terms— Mooney Faces, Generative Adversarial Network, Cross-Tone Integration

1. INTRODUCTION
Mooney faces are special two-tone black and white pictures of faces (Fig.1). Despite the lack of detail, shading, and other cues, human observers can often effortlessly recognize the identity, age, gender, and facial expression of Mooney faces [1, 2, 3]. Even infants can recognize Mooney faces [4, 5, 6]. How such a remarkable perceptual ability is achieved with such sparse information may be key to understanding human face recognition in naturalistic settings, where noise, occlusion, and shadows are common.

Mooney faces may be challenging for both human and computer vision systems because the contours are ambiguous; it is not clear how they can be segmented into parts based on bottom-up image-level information alone. While it is trivial to extract the contours of a shadowy Mooney face, it is not obvious how to interpret or classify these contours. In a two-tone Mooney face, each contour could correspond to a meaningful structure of the face (what is sometimes called an attached shadow) or to a less meaningful cast shadow border. Cast shadow borders depend on the nature and direction of the light source, the shape of the intervening surface, and the surface on which the shadow is cast. Thus, there is skepticism about the amount of usable information from cast shadow borders for human recognition [7]. Attached shadows depend primarily on the structure of the surface, and are therefore useful for recognition. However, parsing the contour into relevant attached shadow segments and irrelevant cast shadow segments requires some prior knowledge about the image. For these reasons, it is thought that Mooney faces require holistic [2] or top-down information for recognition [2, 3], for example by template matching [8].

Face recognition from photos is well studied in computer vision [9, 10], however, current approaches applied out of the box would fail miserably on two-tone Mooney faces. If there were an algorithm for recognizing Mooney faces, it could shed light on human visual processing. It could also shed light on broader issues in computer vision like face recognition in noise, occlusion, and extreme lighting conditions, which are frequently encountered in surveillance videos.

Mooney faces are not well defined; not every binarized face image appears as a Mooney face. Mooney faces are usually constructed by artists or scientists manually, in an ad-hoc manner, requiring selective grayscale image editing and human subjective judgment, a time and labor-consuming process. Such a mechanism for generating Mooney faces does not scale, severely limiting current approaches of studying Mooney faces. Our first goal here is to generate a best possible Mooney face automatically from a face photo. We take a
We develop a deep learning model to automatically generate a plausible Mooney face from a given face photo (Fig.2). We create two-tone Mooney candidates by smoothing and thresholding the face photo with multiple parameter settings. We train a face classifier to determine whether a two-tone image is a Mooney face. We feed the Mooney face candidates into our classifier and the top scorer becomes our automatically generated Mooney face.

The key to our approach is how to train our Mooney face classifier, since we only have a very limited number of Mooney faces found in the literature. We take a two-step approach. We first use large scale FaceScrub dataset [11] and non-face ImageNet dataset [12] to pre-train a grayscale face classifier. We then fine-tune the model with the small number of Mooney faces using various data augmentation techniques.

Grayscale face classifier. We build our collection of positive and negative face image examples to train a grayscale face classifier. Our face photos are 90,000 FaceScrub images of about 500 identities, whereas our non-face photos are 90,000 images uniformly sampled over 1,000 classes of the ImageNet dataset. We extract the model from the triplet network of OpenFace [13] and append it with a two-way softmax layer for representing the output of whether the input image is a face or not. The input image resolution is set to $96 \times 96$. We optimize the network with stochastic gradient descent for 40,000 iterations, with mini-batch size 128, momentum 0.9, and weight decay 0.0005. The learning rate starts at 0.01 and exponentially decays to 0.001.

Mooney face classifier. There are only a few Mooney faces publicly available. Our Mooney face examples come from 48 images of the PICS dataset [14] and our own 80 Mooney faces. It is also known in visual psychology that Mooney faces are not recognizable when the intensity polarity is reversed or the image is inverted. We compose our non-Mooney examples of five types of images: 90,000 binarized ImageNet images, their negative versions (where black pixels become white pixels, and vice versa), the Mooney face inversions, the Mooney face negatives, and their inverted versions.

We fine-tune the previously trained grayscale face classifier for Mooney-face classification. We follow the same hyper-parameter setting for training and testing. We also perform additional data processing for augmentation: Mooney faces are randomly dilated or eroded; ImageNet images are from 48 images of the PICS dataset [14] and our own 80 Mooney faces. It is also known in visual psychology that Mooney faces are not recognizable when the intensity polarity is reversed or the image is inverted. We compose our non-Mooney examples of five types of images: 90,000 binarized ImageNet images, their negative versions (where black pixels become white pixels, and vice versa), the Mooney face inversions, the Mooney face negatives, and their inverted versions.

We fine-tune the previously trained grayscale face classifier for Mooney-face classification. We follow the same hyper-parameter setting for training and testing. We also perform additional data processing for augmentation: Mooney faces are randomly dilated or eroded; ImageNet images are augmented with Photo-to-Mooney data processing.

Photo-to-Mooney data processing. We generate Mooney candidates from a source photo by first smoothing it with a set of $k \times k$ kernels by convolution and then binarizing it with thresholds $t$, where $k \in \{2, 3, 4, 5, 6\}$ and $t \in \{0.4, 0.45, 0.475, 0.5, 0.525, 0.55, 0.6\}$. The pixel intensity is rescaled to the range between 0.4 and 0.6 before thresholding. There could be other alternatives; the goal is simply to generate relatively fewer black-white blobs that tend to be more Mooney-like.
Large-scale Mooney face dataset. We generate 35 candidates for each source face photo from the 35 combinations of $k$ and $t$. Our Mooney face classifier is used to pick out the most Mooney-like candidate. If that candidate has a low score, we reject the face photo. We build our large-scale Mooney face dataset based on Facescrub images. With a high Mooney score threshold to ensure the quality of such automatically generated Mooney faces without human judgement verification, we obtain a Mooney face dataset with 13,460 images of 523 identities, sample results shown in Fig 3.

3. MOONEY-TO-PHOTO PREDICTION

Our large-scale Mooney face dataset with paired source photos and Mooney faces opens up exciting new opportunities. We explore predicting fine-tone faces from two-tone Mooney faces, emulating human vision’s rich face perception from two-tone Mooney faces and enabling computer face recognition in low-bit images from a wide range of conditions.

Related Works. Mooney-to-photo prediction is an interesting and challenging research topic in computer vision, and only a few works have attempted solving the problem. Shashua et al.[15] use linear models to recover faces from two-tone images under several restrictive assumptions, e.g. Lambertian reflection models and 3D face models. Maver et al.[16] build the eigenspace from a set of grayscale faces, use linear models to predict the coefficients from input two-tone images and then reconstruct the grayscale faces. Kemelmacher et al.[17] propose to reconstruct faces from Mooney images with prior knowledge of 3D face models. We approach the problem from a pure data-driven deep learning perspective, translating a Mooney to a face image directly.

Our Patch CGAN Model. Generative Adversarial Networks (GAN) et al.[18] consist of a Discriminator (D) and a Generator (G). G is trained to produce fake data that looks real, whereas D learns to classify data as real or fake. This idea is extended to so-called Conditional GAN (CGAN), where both D and G are conditioned on extra data [19].

Isola et al. applies CGAN to paired input data for cross-modal translation from one type of images to another [20]. Their models use an encoder-decoder net with skip connections for G, which first encodes an input image and then de-
codes the latent representation to generate an output image. They use a convolutional Patch CGAN classifier for D, collecting real/fake scores over 16 × 16 patches. We follow the same Patch CGAN setting to train a CGAN on paired (Mooney,photo) data in order to predict a plausible grayscale face photo from a Mooney image.

**Baseline CCA methods.** Since our D is based on 16 × 16 patches and our G is based on 128 × 128 images, we apply correlation coefficient analysis (CCA) at both batch and full image levels as linear baseline models.

**Experimental Setup.** We split our new Mooney face dataset evenly into train, validation and test sets by identity. We resize the shorter size of each input image to 128 and randomly crop a 128 × 128 patch for training, whereas we only use center crops for testing. Horizontal flipping is used for additional data augmentation. We use ADAM [21] to optimize the network. We set the initial learning rate to 0.0002, momentum 0.5, batch size 1, λ_{L1} 100.

**Cross-Tone Prediction Results.** Fig. 4 shows grayscale face predictions and Table 3 prediction errors. Our CGAN model significantly outperforms CCA models. Our CGAN predictions are strikingly similar to source photos, despite the fact that our network has never seen any faces of the test identities and the inputs are merely two-tone Mooney images. The linear CCA models have trouble generalizing to unseen faces: full-image CCA produces grainy results, whereas patch-wise CCA produces over-smoothed results which lack highlights and shadows and look more like two-tone Mooney faces.

**Are Mooney Faces Special?** We generate alternative two-tone images by intensity thresholding. We binarize each photo by an intensity percentile. Since ground-truth Mooney faces have a mean white pixel percentage of 40%, we choose three percentiles: 30%, 40%, 50%. We also concatenate these binary versions into a color (3 1-bit) image. We follow the same setting to train and test face predictions from such two-tone images. Fig. 5 shows prediction results and Table 3 prediction errors. Mooney faces produce better and non-trivial predictions than simple two-tone images.

**Face Prediction from Very Low-Resolution Surveillance Photos.** We crop 12 × 12 face patches from the VIPeR surveillance dataset[22] and resize them to 128 × 128. We apply our FacePhoto-to-Mooney model to generate its Mooney version, and then apply our Mooney-to-Photo model to predict a grayscale photo. Fig. 6 shows remarkable 128 × 128 faces predicted from 12 × 12 face patches.

**Summary.** We develop deep learning models for Photo-to-Mooney generation and Mooney-to-face prediction. We build the first large-scale paired (face,Mooney) dataset, and learn two-tone Mooney to full-tone face predictions. We show that Mooney images produce better and non-trivial face predictions than intensity thresholded two-tone images. We apply both our models successively to predict faces from extreme low-resolution face images. These results are useful for understanding what makes Mooney faces special and for developing better face recognition methods.
4. REFERENCES


