

## Article

# Effects of Image Quality on the Accuracy Human Pose Estimation and Detection of Eye Lid Opening/Closing Using Openpose and DLib

Run Zhou Ye <sup>1,2,3</sup> , Arun Subramanian <sup>3</sup>, Daniel Diedrich <sup>3</sup>, Heidi Lindroth <sup>3,4</sup> , Brian Pickering <sup>3</sup> and Vitaly Herasevich <sup>3,\*</sup> 

<sup>1</sup> Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 2C4, Canada

<sup>2</sup> Division of Endocrinology, Department of Medicine, Centre de Recherche du CHUS, Sherbrooke, QC J1H 5N4, Canada

<sup>3</sup> Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN 55902, USA

<sup>4</sup> Department of Nursing, Mayo Clinic, Rochester, MN 55902, USA

\* Correspondence: vitaly@mayo.edu



**Citation:** Ye, R.Z.; Subramanian, A.; Diedrich, D.; Lindroth, H.; Pickering, B.; Herasevich, V. Effects of Image Quality on the Accuracy Human Pose Estimation and Detection of Eye Lid Opening/Closing Using Openpose and DLib. *J. Imaging* **2022**, *8*, 330. <https://doi.org/10.3390/jimaging8120330>

Academic Editors: M. Donatello Conte and Pier Luigi Mazzeo

Received: 13 October 2022

Accepted: 15 December 2022

Published: 19 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Objective: The application of computer models in continuous patient activity monitoring using video cameras is complicated by the capture of images of varying qualities due to poor lighting conditions and lower image resolutions. Insufficient literature has assessed the effects of image resolution, color depth, noise level, and low light on the inference of eye opening and closing and body landmarks from digital images. Method: This study systematically assessed the effects of varying image resolutions (from  $100 \times 100$  pixels to  $20 \times 20$  pixels at an interval of 10 pixels), lighting conditions (from 42 to 2 lux with an interval of 2 lux), color-depths (from 16.7 M colors to 8 M, 1 M, 512 K, 216 K, 64 K, 8 K, 1 K, 729, 512, 343, 216, 125, 64, 27, and 8 colors), and noise levels on the accuracy and model performance in eye dimension estimation and body keypoint localization using the Dlib library and OpenPose with images from the Closed Eyes in the Wild and the COCO datasets, as well as photographs of the face captured at different light intensities. Results: The model accuracy and rate of model failure remained acceptable at an image resolution of  $60 \times 60$  pixels, a color depth of 343 colors, a light intensity of 14 lux, and a Gaussian noise level of 4% (i.e., 4% of pixels replaced by Gaussian noise). Conclusions: The Dlib and OpenPose models failed to detect eye dimensions and body keypoints only at low image resolutions, lighting conditions, and color depths. Clinical Impact: Our established baseline threshold values will be useful for future work in the application of computer vision in continuous patient monitoring.

**Keywords:** deep learning; image quality; pose estimation; facial feature extraction

## 1. Introduction

Recent advances in computer vision are being applied in a number of industries including the healthcare sector. Outside of healthcare, numerous algorithms have been applied in autonomous driving [1], facial recognition in airports [2,3], self-service Amazon convenience stores [4], and cybersecurity [5]. In healthcare, those technologies are predominantly used in radiology and other imaging processing. For example, the classification of mammography using convolutional neural networks showed high sensitivity and specificity in detecting breast neoplasms [6–8]. Furthermore, encoder–decoder convolutional networks and cycle-consistent generative adversarial networks have also shown promise in a plethora of image semantic segmentation and translation tasks, including semantic segmentation of organs and tissues using ultrasound [9,10], MRI [11,12], and CT [13–16] images.

Another important area of application of computer vision in the healthcare setting is in continuous hospitalized patients' activity monitoring [17–19]. It has been demonstrated that cameras can be installed in hospital rooms to capture continuous video feeds of

patients; machine-learning algorithms can then be applied to detect body movements, facial expressions, emotions, eyelid and pupil movement, and eye-opening/closing and classify whether the patient is lying down, sitting-up, or ambulating [20]. Assessing patient eyelid opening/closing is particularly important in the context of detecting a patient's level of consciousness (e.g., to calculate the Glasgow coma scale) and recording patient sleep/wake cycles. A priori, this type of system may serve as a detection device for early signs of patient deterioration or risks to patient safety such as falls or delirium without the need for a dedicated observer. Moreover, this new layer of information can also augment traditional patient data to aid clinical decision-making and accurate disease prognostication.

However, there may be several obstacles when applying current computer vision algorithms in the hospital setting [18,21]. First, limitations in camera placement can result in lower relative dimensions of facial and bodily features. Different camera models result in varied or lower video or image resolution and quality, which may be desirable given limited computational resources and the large numbers of patients monitored simultaneously. Furthermore, hospital room lighting is never constant, and lower levels result in darker images with lower color depth and higher noise [22–27]. The low lighting during the night can in part be circumvented by using cameras with larger sensors or infrared imaging [28].

To our knowledge, insufficient literature has assessed the effects of image resolution, color depth, noise level, and low light on the inference of eye opening and closing and body landmarks from digital images.

The aim of the present study is to test the accuracy of commonly used deep-learning models applied to different image resolutions, lighting conditions, color depths, and noise levels to establish baseline threshold values when the quality of the model drops below the accepted level of performance.

These parameters are important to establish for future work in applying computer vision in actual patient monitoring. It can be hypothesized that a degraded image should gradually decrease the model's accuracy up to a certain threshold beyond which the model will fail completely.

Previous literature on human pose estimation and facial landmark detection is first summarized, followed by a description of the methods for decreasing image quality and testing model accuracy. Results of the effects of image resolution, color depth, noise level, and low light are then reported and discussed.

## 2. Related Works

Human pose estimation using deep learning has been the subject of intense research in recent years and has been reviewed in [29]. In general, there are two types of multi-subject pose estimation algorithms. The top-down approach first detects all human subjects in a particular scene and subsequently localizes all keypoints for each given subject. Algorithms that use such a technique include G-RMI [30], Mask-RCNN [31], MSRA [32], CPN [33], and ZoomNet [34]. By combining high- and low-resolution representations through multi-scale fusion while maintaining a high-resolution backbone, HRNet [35] and HigherHRNet [36] achieved excellent keypoint detection results.

In contrast, the bottom-up technique identifies all keypoints first and then assigns each keypoint to an individual subject. Algorithms that employ this method include DeepCut [37], DeeperCut [38], and MultiPoseNet [39]. By introducing part affinity fields, OpenPose became the most popular bottom-up algorithm [40]. The concept of part affinity fields was expanded in PifPaf through the addition of a part intensity field [41].

Facial landmark detection is closely related to pose estimation and has benefited from advancements made in human pose estimation. Algorithms used for facial landmark detection have been previously reviewed in [42]. The earliest algorithms used deformable facial mesh, which has been replaced by an ensemble of regression tree models [43] such as those included in the Dlib open-source library [44]. Since they have very high computation speeds and are easy to implement, these models have become widely used in research. More recently, algorithms used in pose estimation such as HRNet [35] have been adapted

for facial landmark detection [45]. In addition, other newer methods using shape model (e.g., Dense face alignment [46]), heatmap (e.g., style-aggregated network [47], aggregation via separation [48], FAN [49], and MobileFAN [50]), and direct regression (e.g., PFLD [51], deep graph learning [52], and AnchorFace [53]) techniques have been proposed.

### 3. Materials and Methods

#### 3.1. Data

Two hundred images (100 images of humans with eyes open and 100 images with eyes closed) randomly chosen from the Closed Eyes in the Wild Dataset [54] were used to assess model accuracy as image quality was gradually degraded.

To generate out-of-sample images, photographs of the primary author with eyes open and closed were captured using a 13 MP smartphone camera (Moto E XT2052-1, 13 MP, f/2.0, 1/3.1), with the height of the face occupying approximately half of the image height. Images were obtained using three 300-lumen dimmable light sources placed 150 cm in front of the face.

To test the effects of image quality on the accuracy of pose estimation, images from the COCO 2017 [55] validation dataset (<https://cocodataset.org/#overview>, accessed on 23 June 2022) were used. Specifically, images that depict exactly one person (Figure S6) were extracted along with body keypoint annotations (921 images).

#### 3.2. Model Description

Facial landmark recognition was performed using the pretrained model in Dlib v19.24.0 (<http://dlib.net/>, accessed on 20 June 2022). Sixty-eight key facial landmarks were predicted by the model (see Supplementary Figure S5), where points 36 to 41 and 42 to 47 delineate the right and left palpebral fissures, respectively.

Images from the COCO body keypoint dataset were used for pose estimation with the OpenPose (<https://github.com/CMU-Perceptual-Computing-Lab/openpose>, accessed on 12 July 2022) pretrained model v1.7.0 [40].

#### 3.3. Modifications Made

Two hundred images from the Closed Eyes in the Wild Dataset were used to assess model accuracy as image quality was gradually degraded. To generate images of different resolutions (a total of 8000 images), the original images were resized from  $100 \times 100$  pixels to  $20 \times 20$  pixels at an interval of 10 pixels while maintaining the aspect ratio. The image color depth was successively decreased from 16.7 M colors to 8 M, 1 M, 512 K, 216 K, 64 K, 8 K, 1 K, 729, 512, 343, 216, 125, 64, 27, and 8 colors. Gaussian noise was added by replacing randomly chosen pixels with random pixels; noise intensity was changed by varying the probability of replacing a given pixel from 0% to 10% at an interval of 1% and from 10% to 50% at an interval of 10%.

For images captured using the smartphone camera, light intensity at the level of the face was measured using a smartphone light meter application ([https://play.google.com/store/apps/details?id=com.tsang.alan.lightmeter&hl=en\\_CA&gl=US](https://play.google.com/store/apps/details?id=com.tsang.alan.lightmeter&hl=en_CA&gl=US), accessed on 15 July 2022). Images under different lighting conditions were captured by varying light across 21 intensity levels from 42 to 2 lux with an interval of 2 lux. Since the smartphone-captured images had a higher resolution ( $1000 \times 800$  pixels) than images from the Closed Eyes in the Wild Dataset ( $100 \times 100$  pixels), images of different resolutions (a total of 789 images) were generated first by resizing the original images from an image height of 1000 pixels to 100 pixels at an interval of 100 pixels and then from the 100-pixel image height to 10 pixels at an interval of 10 pixels.

For the COCO 2017 keypoint dataset, images of different quality (a total of 44,208 images) were generated. To generate images of different resolutions, the original image width (500 pixels) was decreased to 50 at an interval of 50 pixels and from 50 to 5 pixels at an interval of 5 pixels. Images with different color depths and noise levels were generated with the same quality degradation scheme as outlined above for eye open–closed inference.

### 3.4. Study Procedures

Modifications to the original images and other computations were implemented in Python 3.9. Source codes are available via: <https://figshare.com/s/47540b8b79b16edec831> (accessed on 25 July 2022).

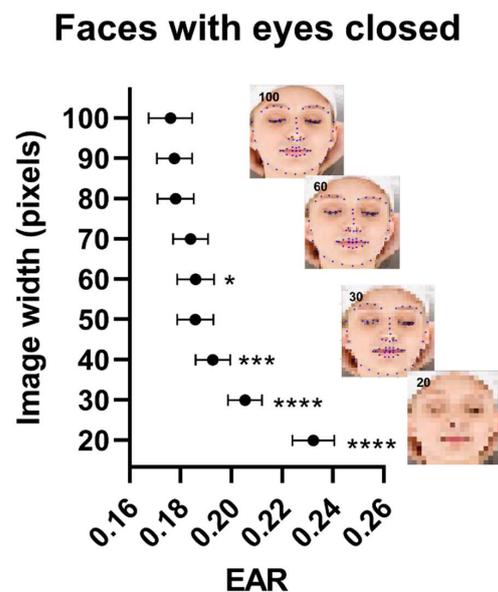
### 3.5. Measurements/Statistics

The opening and closing of the eyes were quantified using the eye aspect ratio as described in [56], which is the ratio of the vertical and horizontal dimensions of the palpebral fissure. Palpebral fissure dimensions were estimated on randomly chosen images (100 images with eyes open and 100 images with eyes closed) from the Closed Eyes in the Wild Dataset. For pose estimation, the mean absolute error (MAE) of the x and y pixel coordinates of the predicted keypoint vs. the ground truth was computed. An average MAE value was calculated for all keypoints. One-way ANOVA with multiple comparisons was carried out using images of the best quality as the comparator. An adjusted *p*-value of less than 0.05 is considered statistically significant. The rate of model failure was computed by dividing the number of images for which the models were unable to detect faces/humans by the total number of images. Statistical calculations were performed in GraphPad Prism 9 and Python 3.9.

## 4. Results

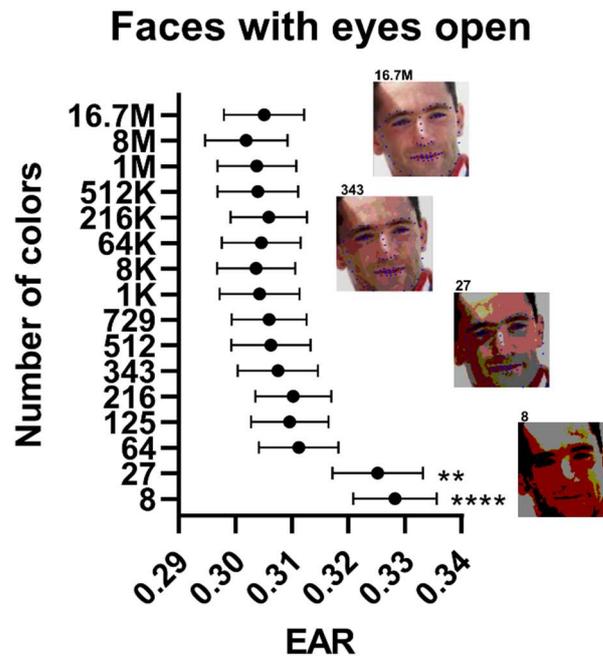
### 4.1. Eye Open-Close Inference

As shown in Figure 1, when image resolution was reduced under 60 pixels  $\times$  60 pixels, model estimates of closed-eye dimensions (EAR of 0.19) deviated from the true dimensions (EAR of 0.18, Figure 1) and the model failed to detect the face and eyes in larger numbers of images (Figure S1D) at 30  $\times$  30 pixels (24%) compared to the baseline (17%). Similar trends can be observed in the open-eye dataset (Figure S2A,D): EAR was 0.30 at the full image resolution (100  $\times$  100 pixels) and deviated to 0.31 when the resolution was decreased to 50  $\times$  50 pixels; missing values increased from 5% at baseline to 10% when the resolution was reduced to 30  $\times$  30 pixels.



**Figure 1.** Palpebral fissure dimensions and model performance in the closed-eyes dataset as a function of image quality using the Closed Eyes in the Wild Dataset. Data points are eye aspect ratio (EAR) estimates as a function of image resolution. Inserts show images at different quality levels with overlaying model prediction. Data points  $\pm$  error represent mean value  $\pm$  SEM. Statistical significance levels were for one-way ANOVA with multiple comparisons using images of the best quality as the comparator. EAR: Eye aspect ratio. \*:  $p < 0.05$ , \*\*\*:  $p < 0.001$ ; \*\*\*\*:  $p < 0.0001$ .

When color depth was reduced from 16.7 M colors to 343 colors, closed-eye dimensions deviated significantly (EAR of 0.18 vs. 0.17 at baseline, Figure S1B). The deviation was highest when the color depth was reduced to 27 colors in the open (EAR of 0.33 vs. 0.30 at baseline, Figure 2) and closed (EAR of 0.19 vs. 0.17 at baseline) eye datasets. Furthermore, the percentage of missing values also increased as the color depth decreased to 343 colors (from 17% to 27% in the closed-eye dataset, Figure S1E, and from 5% to 6% in the open-eye dataset, Figure S2E).

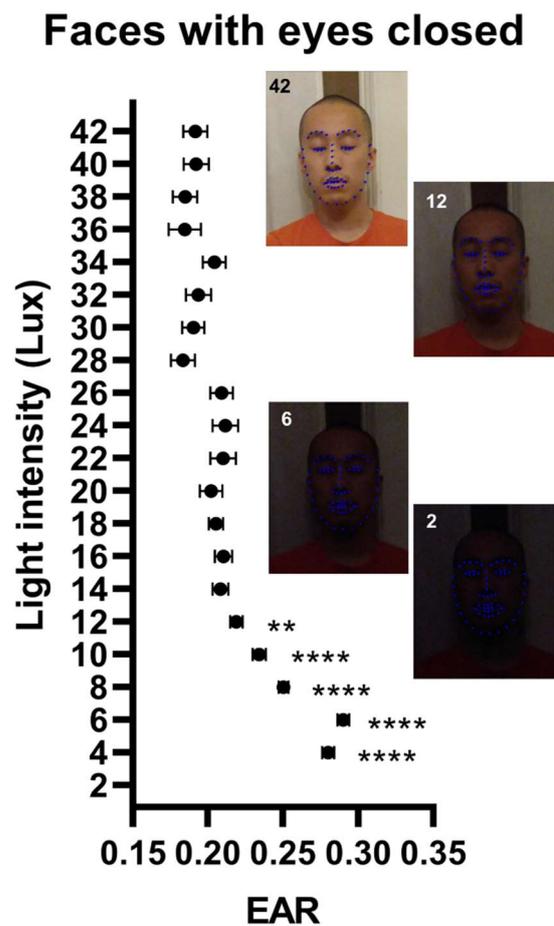


**Figure 2.** Palpebral fissure dimensions and model performance in the open-eyes dataset as a function of image quality using the Closed Eyes in the Wild Dataset. Data points are eye aspect ratio (EAR) estimates as a function of image color depth. Inserts show images at different color depths with overlaying model prediction. Data points  $\pm$  error represent mean value  $\pm$  SEM. Statistical significance levels were for one-way ANOVA with multiple comparisons using images of the best quality as the comparator. EAR: Eye aspect ratio. \*\*:  $p < 0.01$ ; \*\*\*\*:  $p < 0.0001$ .

As shown in Figures S1C and S2C, eye dimension estimates deviated from the true dimensions when 7 to 9% of the original image pixels were replaced by noise (closed-eye EAR of 0.20 with 9% noise vs. 0.17 at baseline and open-eye EAR of 0.32 with 7% noise vs. 0.30 at baseline). However, the percentage of missing values (Figures S1F and S2F) began to increase even when 4% of pixels were replaced by random noise (from 17% to 38% in the closed-eye dataset and from 5% to 10% in the open-eye dataset).

For images with different light intensities, model prediction of palpebral fissure dimension started to deviate from the true dimension as image size was reduced to image heights of 50–70 pixels (EAR of 0.20 at 1000 pixels vs. 0.22 at 70 pixels, Figure S3A, and EAR of 0.34 at 1000 pixels vs. 0.33 at 50 pixels, Figure S3C). Similarly, the number of missing values, i.e., images where the model failed to identify the face and/or both eyes, increased sharply under this image resolution: The percent of missing values increased from 19% at 40 pixels to 95% at 30 pixels in the closed-eye dataset and from 19% at 40 pixels to 76% at 30 pixels in the open-eye dataset.

The model prediction of the palpebral fissure dimension deviated more gradually from the true dimension as the light intensity level decreased under 12 lux (Figures 3 and S3D). At a light intensity of 8 lux, the model was increasingly less capable of correctly identifying the face and both eyes (16% at 42 lux vs. 21% at 8 lux).



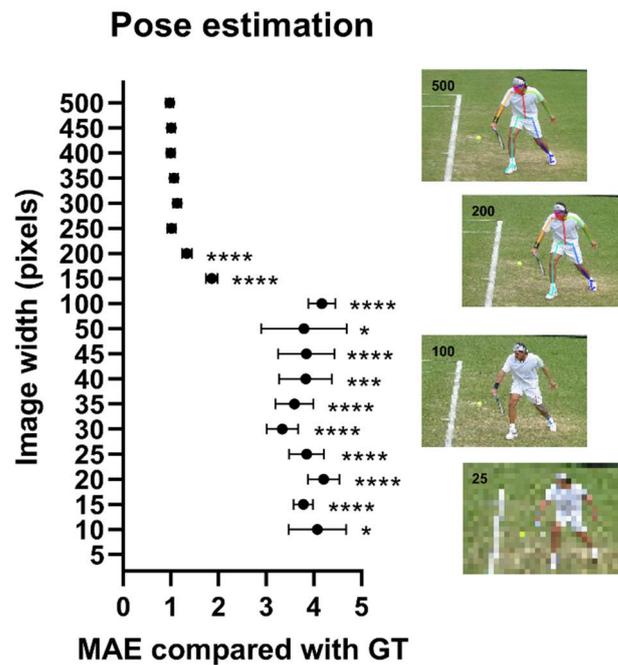
**Figure 3.** Palpebral fissure dimension and model performance as a function of image resolution and light intensity. Data points are the eye aspect ratio (EAR) estimates as a function of image lighting in faces with eyes closed. Inserts show images at different lighting levels with overlaying model prediction. Data points  $\pm$  error represent mean value  $\pm$  SEM. Statistical significance levels were for one-way ANOVA with multiple comparisons using images of the best quality as the comparator. EAR: Eye aspect ratio. \*\*:  $p < 0.01$ ; \*\*\*\*:  $p < 0.0001$ .

#### 4.2. Human Pose Estimation

The prediction accuracy of the model for human poses using the COCO dataset decreased significantly when the image height was reduced to less than 200 of the original 500 pixels (MAE of 1.3 pixels vs. 0.98 pixels, respectively, Figure 4). Since human subjects occupied, on average,  $150 \times 200$  pixels of the original images, this indicates that the model was accurate up to a resolution of  $60 \times 80$  pixels that depict only the human subject. Similarly, the fraction of images where the model was unable to identify the human subject started to increase dramatically beyond this resolution threshold (the percent of missing values increased from 17% at 200 pixels to 84% at 100 pixels, Figure S4D).

When color depth was reduced to values lower than 512 colors, pose estimation began to deviate significantly from the ground truth (MAE of 0.98 pixels at 16.7 M colors vs. 1.12 pixels at 512 colors). The percentage of missing values also increased sharply as color depth was inferior to 343 colors (10% at 16.7 M colors vs. 14% at 343 colors, Figure S4E).

As shown in Figure S4C, the error of pose estimation from ground truth began to rise significantly compared to the baseline when 5% of the original image pixels were replaced by noise (0.97 pixels at baseline vs. 1.17 pixels with 5% noise). Similarly, the percentage of missing values (Figure S4F) started to increase when 4% of pixels were replaced by random noise (10% at baseline vs. 13% with 4% noise).



**Figure 4.** Model performance in the COCO body keypoint dataset as a function of image quality. Data points are mean absolute error values of model prediction as a function of image resolution. Inserts show images at different quality levels with overlaying model prediction. Data points  $\pm$  error represent mean value  $\pm$  SEM. Statistical significance levels were for one-way ANOVA with multiple comparisons using images of the best quality as the comparator. EAR: Eye aspect ratio; GT: Ground truth; MAE: Mean absolute error. \*:  $p < 0.05$ , \*\*\*:  $p < 0.001$ ; \*\*\*\*:  $p < 0.0001$ .

**5. Discussion**

This study systematically tested the effects of image quality on facial feature extraction and human pose estimation using common deep learning models.

For the determination of eye opening and closing with Dlib, the resolution of facial images can be reduced to  $60 \times 60$  pixels without significantly affecting the model estimation of eye dimension. When the color depth of images was lower than 343 colors, eye dimensions estimated by the model began to deviate from the true eye dimensions, and it became increasingly difficult for the model to identify the face. The accuracy of model estimation of eye dimensions began to decrease when 7% of the original image pixels were replaced by noise. Interestingly, even when images of the face were taken under low lighting conditions (14 lux), eye dimensions could still be accurately determined to differentiate between open vs. closed eyes. Under very low lighting (6 lux), the model could still identify the face in most instances.

For human pose estimation using OpenPose, the resolution of regions representing human subjects can be reduced to  $60 \times 80$  pixels without significantly affecting model accuracy or performance. Color depth reduction from 16.7 M to 512 colors resulted in a significant increase in the mean absolute error of model prediction. The addition of more than 4% Gaussian noise also increased model error.

Typically, contemporary convolutional neural networks are trained using images with resolutions greater than a few hundred pixels in width and height. Large image datasets (e.g., Microsoft COCO [55], ImageNet [57], the MPII Human Pose Dataset [58], and the CMU Panoptic Dataset [59]) used for the recognition and pose estimation of human subjects usually contain images with decent resolutions of 300 to 500 pixels in height and width. Images of similar resolutions are also contained in frequently used datasets for facial landmark annotation (e.g., the AFLW Dataset [60] and 300 W [61]) and emotion detection (e.g., AffectNet [62], CK+ [63], and EMOTIC [64]). Furthermore, in medical imaging with

MRI [65,66], PET [65,67], and CT [14,68], deep learning applications are typically trained using images with resolutions ranging from  $128 \times 128$  to  $512 \times 512$  pixels.

Previous studies have investigated the application of pose estimation algorithms in low-resolution images [69,70]. However, insufficient literature has assessed the effects of image resolution, color depth, noise level, and low light on the inference of eye opening and closing and body landmarks from digital images. Therefore, in the present study, the accuracy of commonly used deep-learning models while varying image resolutions, lighting conditions, color depths, and noise levels was tested. This allowed us to establish baseline threshold values for future work applying computer vision in continuous patient monitoring.

Limitations of this work include the use of relatively small datasets of images; therefore, our study may be underpowered to detect changes in model prediction with small decreases in image quality. Furthermore, subjects in the COCO body keypoint dataset do not all occupy the same number of pixels, which may have introduced heterogeneity in model accuracy. Future work may therefore be performed by testing multiple different networks for a given task using larger numbers of images. In addition, variability (e.g., head tilt) exists in the photos captured by the smartphone camera, which may be an obstacle to the reproducibility of the results. In addition, only the OpenPose and DLib models were tested without model finetuning; other newer deep learning models (e.g., Retinaface [71] and Mediapipe [72,73]) should be studied in future works. Future works should also assess the effects of video instead of photo quality on model accuracy.

## 6. Conclusions

In this study, the effects of image quality on facial feature extraction and human pose estimation using the Dlib and OpenPose models were systematically assessed. It is found that, so far, these models only failed to detect eye dimensions and body keypoints at very low image resolutions (the failure rate for eye dimension estimation increased from below 20% to over 70% by decreasing the facial image resolution from  $40 \times 40$  to  $30 \times 30$  pixels), lighting conditions (the failure rate for eye dimension estimation of 16% at 42 lux light intensity vs. 21% at 8 lux), and color depths (failure rate for pose estimation of 10% at 16.7 M colors vs. 14% at 343 colors). Our established baseline threshold values will be essential for future work in the application of computer vision in continuous patient monitoring.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jimaging8120330/s1>, Supplementary material. Supplementary Figure S1. Palpebral fissure dimension and model performance in the closed-eyes dataset as a function of image quality using the Closed Eyes in the Wild Dataset. Supplementary Figure S2. Palpebral fissure dimension and model performance in the open-eyes dataset as a function of image quality using the Closed Eyes in the Wild Dataset. Supplementary Figure S3. Palpebral fissure dimension and model performance as a function of image resolution and light intensity. Supplementary Figure S4. Model performance in the COCO body keypoint dataset as a function of image quality. Supplementary Figure S5. Facial landmarks prediction using Dlib showing an example of model output with the localization of 64 landmarks.

**Author Contributions:** Conceptualization, R.Z.Y. and V.H.; Methodology, R.Z.Y. and V.H.; Software, R.Z.Y.; Validation, R.Z.Y.; Formal analysis, R.Z.Y., A.S., D.D., H.L. and V.H.; Investigation, R.Z.Y., A.S., D.D., H.L., B.P. and V.H.; Data curation, B.P.; Writing—original draft, R.Z.Y. and V.H.; Writing—review & editing, R.Z.Y., A.S., D.D., H.L., B.P. and V.H.; Visualization, R.Z.Y.; Supervision, V.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** R.Z. Ye received funding from the Canadian Institute of Health Research (CIHR Funding Reference Number: 202111FBD-476587-76355).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Modifications to the original images and other computations were implemented in Python 3.9. Source codes are available at [https://drive.google.com/drive/folders/1XnI62\\_b\\_cgTZV1VnNpxmx74OwTFy\\_yCk?usp=sharing](https://drive.google.com/drive/folders/1XnI62_b_cgTZV1VnNpxmx74OwTFy_yCk?usp=sharing) (accessed on 28 July 2022).

**Acknowledgments:** Vitaly Herasevich is the corresponding author of this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yurtsever, E.; Lambert, J.; Carballo, A.; Takeda, K. A survey of autonomous driving: Common practices and emerging technologies. *IEEE Access* **2020**, *8*, 58443–58469. [CrossRef]
2. Balla, P.B.; Jadhao, K. IoT based facial recognition security system. In Proceedings of the 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, India, 5 January 2018; IEEE: New York, NY, USA.
3. Zhang, Z. Technologies raise the effectiveness of airport security control. In Proceedings of the 2019 IEEE 1st International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Kunming, China, 17–19 October 2019; IEEE: New York, NY, USA.
4. Ives, B.; Cossick, K.; Adams, D. Amazon Go: Disrupting retail? *J. Inf. Technol. Teach. Cases* **2019**, *9*, 2–12. [CrossRef]
5. Berman, D.S.; Buczak, A.L.; Chavis, J.S.; Corbett, C.L. A survey of deep learning methods for cyber security. *Information* **2019**, *10*, 122. [CrossRef]
6. Shen, L.; Margolies, L.R.; Rothstein, J.H.; Fluder, E.; McBride, R.; Sieh, W. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Sci. Rep.* **2019**, *9*, 12495. [CrossRef]
7. Yala, A.; Lehman, C.; Schuster, T.; Portnoi, T.; Barzilay, R. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. *Radiology* **2019**, *292*, 60–66. [CrossRef]
8. Becker, A.S.; Marcon, M.; Ghafoor, S.; Wurnig, M.C.; Frauenfelder, T.; Boss, A. Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer. *Investig. Radiol.* **2017**, *52*, 434–440. [CrossRef]
9. Milletari, F.; Ahmadi, S.-A.; Kroll, C.; Plate, A.; Rozanski, V.; Maiostre, J.; Levin, J.; Dietrich, O.; Ertl-Wagner, B.; Bötzel, K. Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound. *Comput. Vis. Image Underst.* **2017**, *164*, 92–102. [CrossRef]
10. Liu, S.; Wang, Y.; Yang, X.; Lei, B.; Liu, L.; Li, S.X.; Ni, D.; Wang, T. Deep learning in medical ultrasound analysis: A review. *Engineering* **2019**, *5*, 261–275. [CrossRef]
11. Akkus, Z.; Galimzianova, A.; Hoogi, A.; Rubin, D.L.; Erickson, B.J. Deep learning for brain MRI segmentation: State of the art and future directions. *J. Digit. Imaging* **2017**, *30*, 449–459. [CrossRef]
12. Avendi, M.R.; Kheradvar, A.; Jafarkhani, H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med. Image Anal.* **2016**, *30*, 108–119. [CrossRef] [PubMed]
13. Gibson, E.; Giganti, F.; Hu, Y.; Bonmati, E.; Bandula, S.; Gurusamy, K.; Davidson, B.; Pereira, S.P.; Clarkson, M.J.; Barratt, D.C. Automatic multi-organ segmentation on abdominal CT with dense v-networks. *IEEE Trans. Med. Imaging* **2018**, *37*, 1822–1834. [CrossRef] [PubMed]
14. Weston, A.D.; Korfiatis, P.; Kline, T.L.; Philbrick, K.A.; Kostandy, P.; Sakinis, T.; Sugimoto, M.; Takahashi, N.; Erickson, B.J. Automated abdominal segmentation of CT scans for body composition analysis using deep learning. *Radiology* **2019**, *290*, 669–679. [CrossRef] [PubMed]
15. Ye, R.Z.; Noll, C.; Richard, G.; Lepage, M.; Turcotte, E.E.; Carpentier, A.C. DeepImageTranslator: A free, user-friendly graphical interface for image translation using deep-learning and its applications in 3D CT image analysis. *SLAS Technol.* **2022**, *27*, 76–84. [CrossRef] [PubMed]
16. Ye, R.Z.; Montastier, E.; Noll, C.; Frisch, F.; Fortin, M.; Bouffard, L.; Phoenix, S.; Guerin, B.; Turcotte, E.E.; Carpentier, A.C. Total Postprandial Hepatic Nonesterified and Dietary Fatty Acid Uptake Is Increased and Insufficiently Curbed by Adipose Tissue Fatty Acid Trapping in Prediabetes With Overweight. *Diabetes* **2022**, *71*, 1891–1901. [CrossRef]
17. Magi, N.; Prasad, B. Activity Monitoring for ICU Patients Using Deep Learning and Image Processing. *SN Comput. Sci.* **2020**, *1*, 123. [CrossRef]
18. Davoudi, A.; Malhotra, K.R.; Shickel, B.; Siegel, S.; Williams, S.; Ruppert, M.; Bihorac, E.; Ozrazgat-Baslanti, T.; Tighe, P.J.; Bihorac, A. The intelligent ICU pilot study: Using artificial intelligence technology for autonomous patient monitoring. *arXiv* **2018**, arXiv:1804.10201.
19. Ahmed, I.; Jeon, G.; Piccialli, F. A deep-learning-based smart healthcare system for patient’s discomfort detection at the edge of Internet of things. *IEEE Internet Things J.* **2021**, *8*, 10318–10326. [CrossRef]
20. Yeung, S.; Rinaldo, F.; Jopling, J.; Liu, B.; Mehra, R.; Downing, N.L.; Guo, M.; Bianconi, G.M.; Alahi, A.; Lee, J.; et al. A computer vision system for deep learning-based detection of patient mobilization activities in the ICU. *NPJ Digit. Med.* **2019**, *2*, 11. [CrossRef]
21. Davoudi, A.; Malhotra, K.R.; Shickel, B.; Siegel, S.; Williams, S.; Ruppert, M.; Bihorac, E.; Ozrazgat-Baslanti, T.; Tighe, P.J.; Bihorac, A. Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. *Sci. Rep.* **2019**, *9*, 8020. [CrossRef]

22. Rahim, A.; Maqbool, A.; Rana, T. Monitoring social distancing under various low light conditions with deep learning and a single motionless time of flight camera. *PLoS ONE* **2021**, *16*, e0247440. [[CrossRef](#)]
23. Ren, W.; Liu, S.; Ma, L.; Xu, Q.; Xu, X.; Cao, X.; Du, J.; Yang, M.-H. Low-light image enhancement via a deep hybrid network. *IEEE Trans. Image Process.* **2019**, *28*, 4364–4375. [[CrossRef](#)] [[PubMed](#)]
24. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [[CrossRef](#)] [[PubMed](#)]
25. McCunn, L.J.; Safranek, S.; Wilkerson, A.; Davis, R.G. Lighting control in patient rooms: Understanding nurses' perceptions of hospital lighting using qualitative methods. *HERD Health Environ. Res. Des. J.* **2021**, *14*, 204–218. [[CrossRef](#)] [[PubMed](#)]
26. Bernhofer, E.I.; Higgins, P.A.; Daly, B.J.; Burant, C.J.; Hornick, T.R. Hospital lighting and its association with sleep, mood and pain in medical inpatients. *J. Adv. Nurs.* **2014**, *70*, 1164–1173. [[CrossRef](#)] [[PubMed](#)]
27. Leccese, F.; Montagnani, C.; Iaia, S.; Rocca, M.; Salvadori, G. Quality of lighting in hospital environments: A wide survey through in situ measurements. *J. Light Vis. Environ.* **2016**, *40*, 52–65. [[CrossRef](#)]
28. Ring, E.; Ammer, K. The technique of infrared imaging in medicine. In *Infrared Imaging: A Casebook in Clinical Medicine*; IOP Publishing: Bristol, UK, 2015.
29. Liu, W.; Mei, T. Recent Advances of Monocular 2D and 3D Human Pose Estimation: A Deep Learning Perspective. *ACM Comput. Surv. (CSUR)* **2022**, *55*, 80. [[CrossRef](#)]
30. Papandreou, G.; Zhu, T.; Kanazawa, N.; Toshev, A.; Tompson, J.; Bregler, C.; Murphy, K. Towards accurate multi-person pose estimation in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
31. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
32. Xiao, B.; Wu, H.; Wei, Y. Simple baselines for human pose estimation and tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
33. Chen, Y.; Wang, Z.; Peng, Y.; Zhang, Z.; Yu, G.; Sun, J. Cascaded pyramid network for multi-person pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
34. Jin, S.; Xu, L.; Xu, J.; Wang, C.; Liu, W.; Qian, C.; Ouyang, W.; Luo, P. Whole-body human pose estimation in the wild. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2020.
35. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
36. Cheng, B.; Xiao, B.; Wang, J.; Shi, H.; Huang, T.S.; Zhang, L. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
37. Pishchulin, L.; Insafutdinov, E.; Tang, S.; Andres, B.; Andriluka, M.; Gehler, P.V.; Schiele, B. Deepcut: Joint subset partition and labeling for multi person pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
38. Insafutdinov, E.; Pishchulin, L.; Andres, B.; Andriluka, M.; Schiele, B. Deepcut: A deeper, stronger, and faster multi-person pose estimation model. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2016.
39. Kocabas, M.; Karagoz, S.; Akbas, E. Multiposenet: Fast multi-person pose estimation using pose residual network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
40. Cao, Z.; Simon, T.; Wei, S.-E.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
41. Kreiss, S.; Bertoni, L.; Alahi, A. Pifpaf: Composite fields for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA., 15–20 June 2019.
42. Khabaralax, K.; Koriashkina, L. Fast facial landmark detection and applications: A survey. *arXiv* **2021**, arXiv:2101.10808. [[CrossRef](#)]
43. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
44. King, D.E. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.
45. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X. Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3349–3364. [[CrossRef](#)]
46. Liu, Y.; Jourabloo, A.; Ren, W.; Liu, X. Dense face alignment. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017.
47. Dong, X.; Yan, Y.; Ouyang, W.; Yang, Y. Style aggregated network for facial landmark detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
48. Qian, S.; Sun, K.; Wu, W.; Qian, C.; Jia, J. Aggregation via separation: Boosting facial landmark detector with semi-supervised style translation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.
49. Bulat, A.; Tzimiropoulos, G. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.

50. Zhao, Y.; Liu, Y.; Shen, C.; Gao, Y.; Xiong, S. Mobilefan: Transferring deep hidden representation for face alignment. *Pattern Recognit.* **2020**, *100*, 107114. [[CrossRef](#)]
51. Guo, X.; Li, S.; Yu, J.; Zhang, J.; Ma, J.; Ma, L.; Liu, W.; Ling, H. PFLD: A practical facial landmark detector. *arXiv* **2019**, arXiv:1902.10859.
52. Li, W.; Lu, Y.; Zheng, K.; Liao, H.; Lin, C.; Luo, J.; Cheng, C.-T.; Xiao, J.; Lu, L.; Kuo, C.-F. Structured landmark detection via topology-adapting deep graph learning. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2020.
53. Xu, Z.; Li, B.; Yuan, Y.; Geng, M. AnchorFace: An anchor-based facial landmark detector across large poses. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, Canada, 2–9 February 2021.
54. Song, F.; Tan, X.; Liu, X.; Chen, S. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recognit.* **2014**, *47*, 2825–2838. [[CrossRef](#)]
55. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2014.
56. Soukupová, T.; Cech, J. Real-Time Eye Blink Detection using Facial Landmarks. In Proceedings of the 21st Computer Vision Winter Workshop, Rimske Toplice, Slovenia, 3–5 February 2016.
57. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: New York, NY, USA.
58. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2d human pose estimation: New benchmark and state of the art analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
59. Joo, H.; Liu, H.; Tan, L.; Gui, L.; Nabbe, B.; Matthews, I.; Kanade, T.; Nobuhara, S.; Sheikh, Y. Panoptic studio: A massively multiview system for social motion capture. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.
60. Koestinger, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; IEEE: New York, NY, USA.
61. Sagonas, C.; Tzimiropoulos, G.; Zafeiriou, S.; Pantic, M. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, NSW, Australia, 2–8 December 2013.
62. Mollahosseini, A.; Hasani, B.; Mahoor, M.H. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Trans. Affect. Comput.* **2017**, *10*, 18–31. [[CrossRef](#)]
63. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; IEEE: New York, NY, USA.
64. Kosti, R.; Alvarez, J.M.; Recasens, A.; Lapedriza, A. Context based emotion recognition using emotic dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2755–2766. [[CrossRef](#)]
65. Liu, F.; Jang, H.; Kijowski, R.; Bradshaw, T.; McMillan, A.B. Deep Learning MR Imaging-based Attenuation Correction for PET/MR Imaging. *Radiology* **2018**, *286*, 676–684. [[CrossRef](#)]
66. Herent, P.; Schmauch, B.; Jehanno, P.; Dehaene, O.; Saillard, C.; Balleyguier, C.; Arfi-Rouche, J.; Jegou, S. Detection and characterization of MRI breast lesions using deep learning. *Diagn. Interv. Imaging* **2019**, *100*, 219–225. [[CrossRef](#)]
67. Ye, E.Z.; Ye, E.H.; Ye, R.Z. DeepImageTranslator V2: Analysis of multimodal medical images using semantic segmentation maps generated through deep learning. *HighTech Innov. J.* **2022**, *3*, 3. [[CrossRef](#)]
68. Koitka, S.; Kroll, L.; Malamutmann, E.; Oezcelik, A.; Nensa, F. Fully automated body composition analysis in routine CT imaging using 3D semantic segmentation convolutional neural networks. *Eur. Radiol.* **2021**, *31*, 1795–1804. [[CrossRef](#)]
69. Wang, C.; Zhang, F.; Zhu, X.; Ge, S.S. Low-resolution human pose estimation. *Pattern Recognit.* **2022**, *126*, 108579. [[CrossRef](#)]
70. Chi, C.; Zhang, D.; Zhu, Z.; Wang, X.; Lee, D.-J. Human pose estimation for low-resolution image using 1-D heatmaps and offset regression. *Multimed. Tools Appl.* **2022**, 1–19. [[CrossRef](#)]
71. Deng, J.; Guo, J.; Ververas, E.; Kotsia, I.; Zafeiriou, S. Retinaface: Single-shot multi-level face localisation in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
72. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.-L.; Yong, M.G.; Lee, J. Mediapipe: A framework for building perception pipelines. *arXiv* **2019**, arXiv:1906.08172.
73. Bazarevsky, V.; Kartynnik, Y.; Vakunov, A.; Raveendran, K.; Grundmann, M. Blazeface: Sub-millisecond neural face detection on mobile gpus. *arXiv* **2019**, arXiv:1907.05047.