

# Seatbelt Segmentation Using Synthetic Images

Sumit Jha

*The University of Texas at Dallas*

Richardson, USA

sumit.jha@utdallas.edu

Isaac Brooks

*The University of Texas at Dallas*

Richardson, USA

isaac.brooks@utdallas.edu

Soumitry J. Ray

*EdgeTensor*

Dallas, USA

soumitryjray@edgetensor.com

Rajesh Narasimha

*EdgeTensor*

Dallas, USA

rajeshnarasimha@edgetensor.com

Naofal Al-Dhahir

*The University of Texas at Dallas*

Richardson, USA

aldhahir@utdallas.edu

Carlos Busso

*The University of Texas at Dallas*

Richardson, USA

busso@utdallas.edu

**Abstract**—Recent advancement in deep learning has led to an increased interest in image processing and computer vision applications for driver monitoring systems. One of the applications where these techniques can be useful is in segmenting and tracking seatbelts. A seatbelt is an important safety feature in the vehicle that if properly used can save lives. Efficient segmentation of the seatbelts in an image provides important information about the correct use of seatbelts. The challenge in developing deep learning algorithms for seatbelt detection and segmentation is the manual annotations required for this task, which is cumbersome. This paper explores a novel formulation to efficiently train a seatbelt model with minimal supervision. We exploit the textureless and shape characteristics of the seatbelts to programmatically synthesize images. Our proposed method synthetically creates images that resemble seatbelt patterns. After training a model exclusively with synthetic images, we iteratively fine-tune it using naturalistic images extracted from online video-sharing websites. The labels for these images are pseudo-labels assigned by the model to confident predictions. Fine-tuning helps adapt the model to better work on real naturalistic images, improving the performance of the system. We obtain an F1-score of 0.55 in segmenting the seatbelt with this approach. We also experiment with fine-tuning the model with a small number of naturalistic images with annotated labels. After pretraining on synthetic samples and pseudo-labeled naturalistic images, we achieve an F1-score of 0.67 using only 200 annotated images.

**Index Terms**—Seatbelt detection, training with synthetic data, in-vehicle safety system.

## I. INTRODUCTION

Advances in computer vision have led to better vision-based solutions for monitoring driver behaviors. Recent advancements have explored various areas relevant for in-vehicle systems including detecting visual attention [1]–[5], tracking upper body activity [6], [7], inferring visual and cognitive distractions [8], [9] and recognizing emotions [10]. These solutions help a smart system to understand the state of the driver. The use of seatbelts is one of the important aspects that are relevant for driver safety inside the car [11], [12]. While there are mechanisms in modern cars to check if a seatbelt is latched, it is easy to trick a system by just connecting the belt behind the driver. Additionally, these systems are not able to detect if the driver is correctly wearing the seatbelt, which is

also important [13]. A promising alternative to determine if a driver or a passenger is wearing a seatbelt is using computer vision solutions. Studies have attempted to detect seatbelts with image-based solutions from either inside or outside the vehicle [14]–[17]. Some of these methods were intended as a solution for traffic enforcement systems.

Our objective is to segment the seatbelt pixels from the image that captures the driver’s upper body. The information provided by the proposed model can be valuable to assess not only whether the driver is using the seatbelt, but also whether she/he is properly using the seatbelt. Developing a robust seatbelt detector poses the following challenges: (a) camera pose invariance, (b) occlusion, (c) image modality (color or infrared), and (d) environmental conditions. Therefore, any data used for training a machine learning model, such as *convolutional neural networks* (CNNs), has to capture these variations. The straightforward approach to address this problem involves collecting and annotating large amounts of images. The annotation involves assigning to each image a binary mask that identifies the placement of the seatbelt (Fig. 1). This approach is expensive and time-demanding. It often overfits the training data. It is also susceptible to environmental variations that may not be well represented in the data such as color variations, twists, or unexpected positions of the seatbelt.

We propose a seatbelt segmentation method relying on minimal supervision to circumvent challenges associated with seatbelt detection and segmentation by following a synthesis approach. We propose to create synthetic images resembling seatbelt patterns. We exploit the textureless and shape characteristics of the seatbelts to programmatically synthesize images. We create these stripes using cubic splines with varied thicknesses, adding random colors and random textures. We also add random backgrounds to the images to simulate varied clothing textures that a driver/passenger may be wearing. This approach does not require manual annotation and can be scaled to create a large training set. Using the location of the seatbelts in the image, we can construct the ground truth for these synthetic images. Instead of using binary masks, we propose to use a Gaussian mask as heatmaps, so seatbelts do not have hard boundaries, resembling patterns observed in natural conditions.

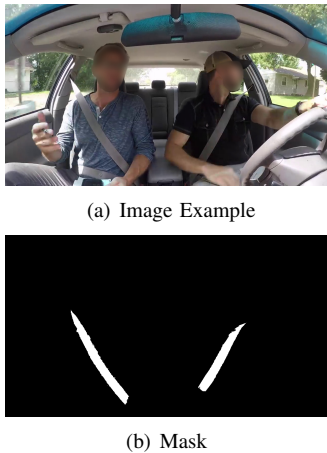


Fig. 1: Example of a seatbelt annotation. A binary mask is needed to identify the pixels included in the seatbelt.

Using the generated images, we train a segmentation model based on CNNs, using the U-net model [18] with a single channel. Since the patterns in the training data have a similar appearance to seatbelts, the convolution filters are expected to learn weights that activate when a similar pattern is observed. Hence, we expect to get high activation in the seatbelt regions when the model processes real images showing the driver’s upper body. Using only the synthetic data, we obtain an F1-score of 0.50 in predicting seatbelts. After training a base model exclusively with synthetic data, we add realistic images extracted from video-sharing websites to fine-tune our model with pseudo-labels using an iterative approach. First, the current model is used to predict seatbelts in real images. Then, the images with confident seatbelt predictions with areas above a given threshold are included in the training set to fine-tune this model. This process is repeated multiple times, updating the model in each iteration. The model achieves an F1-score of 0.55 with just four iterations. This result is promising, since this unsupervised approach does not require manual annotation of images. Finally, we demonstrate that the proposed approach trained with synthetic images can also be used as a pre-trained model when limited labeled images are available to train the model. With only 200 images, we are able to significantly improve the model with an F1-score of 0.67.

## II. RELATED WORK

### A. Seatbelt Detection and Segmentation

Studies have attempted to detect seatbelts from images taken from outside the car using road cameras. This is particularly important for automatic seatbelt enforcement strategies. Zhou et al. [14] extracted important edge features from the salient gradient map of an image containing the driver’s upper body from surveillance images. These features were then used to train a neural network based on the *radial basis function* (RBF), which detected if the seatbelts were present in the image. Guo et al. [15] extracted the direction information from the *hue, saturation, value* (HSV) color space. They performed edge detection followed by the Hough line detection to identify

different lines in an image. They used the lengths and the number of lines in the image to define if an image contained a seatbelt. Zhou et al. [16] trained a model based on AlexNet [19] with batch normalization [20]. This model had a binary output that predicted if a subject was wearing a seatbelt. Elihos et al. [17] evaluated different algorithms for seatbelt detection using traffic images taken with a camera outside the car. They reported that the best model was achieved by a system using the *single shot multi box detector* (SSD) [21].

Studies have also addressed the detection and segmentation of seatbelts using images taken inside the car. Chun et al. [22] train a *feature pyramid network* (FPN) with different heads to estimate the body pose of the driver or passenger. The model also included the task of seatbelt segmentation. They trained the seatbelt using a manually annotated dataset recorded from 100 different subjects in 50 different sessions.

### B. Use of Synthetic Training Data

Some approaches have also explored the use of synthetic data to train models. For example, Ward et al. [23] generated images of ships using a Unity game engine to train a model for ship classification from top view images. They adapted the model to real-world data by fine-tuning the model using a small real-world data set. Similarly, Jaderberg et al. [24] used synthetic data for text recognition in natural scenes (e.g., billboards). To the best knowledge of the authors, this paper is the first study that uses an approach based on synthetic data generation for seatbelt segmentation.

### C. Pseudo-Labels and Semi-Supervised Learning

Semi-supervised learning (SSL) is a learning approach that is useful for problems with a low amount of labeled data. In SSL, the aim is to formulate a training pipeline that requires only a small amount of labeled data. This data-efficient approach is appealing for the problem of image segmentation, since pixel-wise annotations are expensive to annotate. Pseudo-labeling is an iterative SSL approach that involves using model predictions in one iteration as labels in the next iteration [25]. Some implementations, such as the one described by Zou et al. [26], use a consistency loss and image augmentation strategy to improve performance in low-data scenarios.

## III. PROPOSED APPROACH

Our proposed approach generates synthetic data to train a seatbelt segmentation model. Instead of manually annotating a large dataset to train our models, we rely on synthetic data to learn the patterns associated with the seatbelt. Our approach does not require labeled data, so our solution is an appealing approach to address this problem. The available information from these patterns is used to train a CNN-based semantic segmentation model. Furthermore, the models are fine-tuned with real unlabeled data using pseudo-labels generated from the learned model. This section explains the generation process to create synthetic data (Sec. III-A), the proposed DNN model (Sec. III-B), and the fine-tuning approach using pseudo-labels (Sec. III-C).

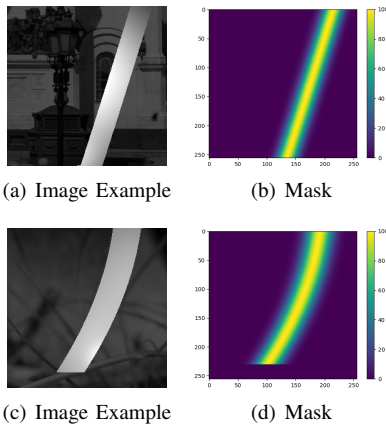


Fig. 2: Synthetic strip-like patterns resembling seatbelts created to train our models. The figure also shows the Gaussian masks associated with the synthetic images.

### A. Synthetic Data Generation

Our approach creates synthetic data resembling seatbelt patterns. We exploit the textureless and shape characteristics of a seatbelt to programmatically synthesize images. Our approach relies on third-order polynomial approximation to generate stripes that resemble seatbelts. A curved line is drawn by performing a spline interpolation between three arbitrary points. Two points are selected in the two opposite corners and the third point is chosen in the lower half of the frame. Using this line as a reference a strip of uniform thickness is drawn. The strip goes from the top right corner to the bottom left corner, or from the top left corner to the bottom right corner. The strips are parametrized with thickness, curvature, orientation angle, and position on the image. Twists are also added in a few examples. This approach is flexible where we can easily change the thickness of the stripes. We change the illumination to simulate different conditions. The background of the seatbelts can significantly vary depending on the clothing of the driver or passengers in the car. Therefore, we incorporate this variability by adding the generated stripes in images with random backgrounds. This approach allows our deep neural network to focus on seatbelt patterns, regardless of the clothes of the driver or passengers. By incorporating randomness in the background, we discourage the model from learning any particular background pattern for a seatbelt.

An important step in our model is to simulate conditions often observed with seatbelts, including different textures and occlusions. The images generated with our approach lack (a) noise associated with the image formation process in an image sensor, and (b) variation in different environmental illumination conditions. Therefore, we add noise drawn from different models to simulate camera noise and environmental changes. Some examples of such camera noises are image sharpening, camera gain noise, and image blur. Similarly, examples of environmental changes are specularity and occlusion. We can recreate these challenges in the created seatbelts by adding random occlusions, removing portions of the strips, and changing the location of the specular reflection.

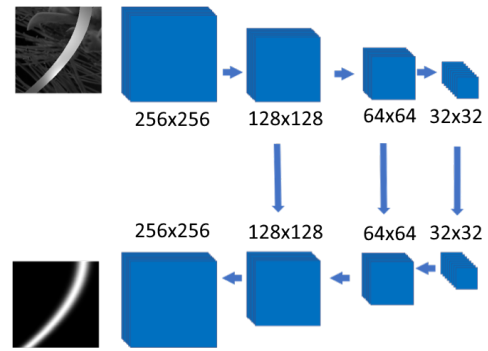


Fig. 3: Proposed U-Net model to segment seatbelts using CNNs. The input is an image with the seatbelts and the output is the predicted masks.

The proposed approach of generating synthesized data does not require manual annotation. Using the location of the belt in the image, we can construct the ground truth for these synthetic images. Instead of using binary masks, we use a Gaussian mask as heatmaps, so seatbelts do not have hard boundaries, resembling challenges observed in natural conditions. Some examples of the synthetic seatbelts with the Gaussian masks are shown in Figure 2. An advantage of this approach is that it can be easily scaled to create a large training set. Using this method, we create a dataset with 500,000 grayscale images with varying seatbelt patterns on random background images taken from the places365 dataset [27]. The images are grayscale to reduce both the size of the dataset and the computational intensity required for training the model. The size of each image is  $256 \times 256$ .

### B. Seatbelt Segmentation Model

The dataset with the synthetic images is used to train the segmentation models to locate these stripe-like patterns in the image. Using the generated images, we train a segmentation model based on CNNs. We use a U-Net model [18] with a single channel input and a single channel output. Figure 3 shows the proposed model. This model consists of a sequence of convolution with maxpooling followed by transposed convolution with half-stride that acts as upsampling. Each convolution block consists of two  $3 \times 3$  convolutions. We use lateral connections to transfer information from the downsampling layers to the corresponding upsampling layers.

The input of the model is the synthetically created seatbelt patterns and the output is a Gaussian mask around the seatbelt region. The model is optimized with the *mean absolute error* (MAE) loss between the predicted mask and the ground truth. Since the patterns in the training data have a similar appearance to seatbelts, the convolution filters are expected to learn weights that activate when a similar pattern is observed. Hence, we expect to get high activation in the seatbelt regions when the model processes real images. We train the base model for 50 epochs with the synthetic data. This model is then fine-tuned with unlabeled data from naturalistic driving images using pseudo-labels (Sec. III-C). We select the model created on the last epoch.

### C. Fine-Tuning with Pseudo-Labels

We fine-tune our final models after training the models exclusively with synthetic images. Our proposed approach is inspired by deep-learning frameworks trained with pseudo-labels [25]. This approach helps the model to learn information pertaining to the driving environment. Iteration zero starts with the model trained with the synthetic images. Then, we use real data with images inside the car. Section IV-A presents the database used for this purpose. We use a total of 5,828 images for fine-tuning the models with pseudo-labels in each iteration. While we have labels for these images, we do not use them for our unsupervised approach. Instead, we select the images with high-confidence predictions and use these predictions as labels for the next iteration of training. We set a threshold to select images only if the seatbelt prediction has a positive area of at least 1,000 pixels, avoiding training our model with unreliable data with small activated regions. We rely on data augmentation by creating multiple variations of these images using the `imgaug` tool [28]. This approach ensures consistency in the model performance against variations in the image. This approach of training the models with real data using pseudo-labels can be applied multiple times. We follow four iterations of fine-tuning. After each iteration, new data with pseudo-labels is selected and added to the database to train the model.

## IV. EXPERIMENTAL SETTING

### A. Database

We require annotations to evaluate the performance of our model. We also need real unlabeled images to implement the pseudo-label approach described in Section III-C. Therefore, we collect data from various video-sharing websites, where a camera is placed inside the vehicle cabin. These videos are obtained from shows and various driving vlogs. The videos are first automatically clustered into different views. Then, we select images captured from views where the driver’s upper body and seatbelt are visible. This is a challenging data set, given the variability across the quality and placement of the cameras. Figure 4 illustrates the process implemented in each image. For each of these images, we run OpenPose [29] to find the shoulder location of each subject. Then, the images are rotated and cropped around each subject. These square images are rotated, resized, and zero-padded to a resolution of  $256 \times 256$ .

We manually annotate a subset of this data set with binary masks for seatbelts similar to the masks shown in Figure 1(b). We select a total of 5,164 images taken from multiple videos at various angles, which are annotated with masks for seatbelts. Since each image may have more than one subject, the total number of seatbelt samples is greater than the original number of images. The total number of annotated seatbelts is 8,129. These seatbelt images are split into two sets. The first set has 5,828 segmented seatbelts, which are used to train our unsupervised approach. In Section V, we also demonstrate that our approach can serve as a pre-trained model for training with limited data. We also use this set for this evaluation. The

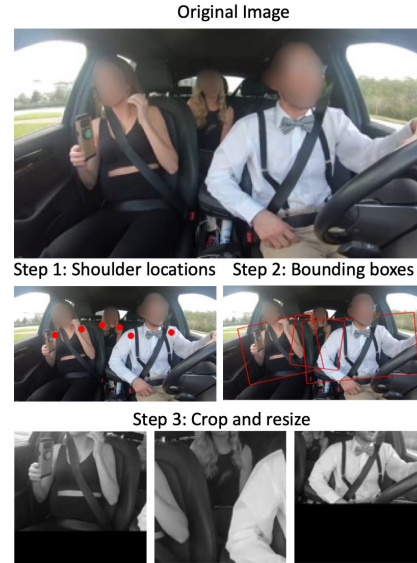


Fig. 4: Pipeline to process real driving images. OpenPose is used to extract shoulder location. The shoulder is the reference to rotate and crop images pertaining to each person.

second set includes the remaining 2,301 segmented seatbelts, which are reserved to evaluate the performance of our model.

## V. EVALUATION RESULTS

We are interested in the correct segmentation of seatbelts, not just a binary classifier that determines whether a driver or passenger is using a seatbelt. Our performance metric reflects this goal. The implementation of our U-Net model creates a heatmap for the detected seatbelt. Our first step is to transform this heatmap into a binary mask by thresholding the value. This step is implemented with a score threshold set to 0.2. A prediction is counted as a *true positive* (TP) if the overlap between the true map and the predicted map is greater than 20% of the area of the true seatbelt. Otherwise, it is considered as a *false negative* (FN). Similarly, a *false positive* (FP) is counted if a false prediction of a size similar to the seatbelt is present (80% in area compared to the true seatbelt). Using these metrics, we estimate the precision rate, recall rate, and the F1-score of the predictions.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1-score} = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{TP + \frac{1}{2}(FP + FN)} \quad (3)$$

Table I shows the precision, recall, and the F1-score of the predictions at different steps. Iteration 0 shows the result of the base model trained only with the synthetic data. Using only synthetic data, the model is able to achieve a F1-score of 0.51. The results for iterations 1, 2, and 3 correspond to the results for the model trained by adding naturalistic driving data with pseudo-labels. We observe a consistent improvement

TABLE I: Performance of the seatbelt segmentation model at different iterations. The iteration 0 corresponds to the base model trained with synthetic images. In iterations 1, 2 and 3, naturalistic driving images are added with pseudo-labels.

Ite.	Precision	Recall	F1-score	IoU (%)
0	0.58	0.43	0.50	19.8
1	0.58	0.48	0.53	21.7
2	0.60	0.50	0.54	22.6
3	0.60	0.51	0.55	22.8
4	0.61	0.51	0.55	23.0

in performance as we introduce naturalistic data with pseudo-labels. As we keep adding more data during the training of the model, we observe steady improvement in the model. We observe false positives caused by objects present in the image with similar shapes to seatbelts, such as arms and steering wheel (Fig. 6). The pseudo-labels may include some of these errors, affecting the model’s effectiveness. Despite these false positives, the F1-score increases with each iteration.

Our model can also be used as a pre-trained model to be fine-tuned with limited labeled data. This evaluation can also determine the performance gain achieved by adding a small amount of annotated naturalistic samples to a pre-trained model. The evaluation considers two pre-trained models: the model exclusively trained with synthetic data (*Model-Iteration 0*), and the model from the fourth iteration trained with real images using pseudo-labels (*Model-Iteration-4*). Then, we randomly select different amounts of training data to fine-tune each base model. We perform this fine-tuning with 25, 50, 100, 150, and 200 training samples, using the same samples for both base models. We also fine-tune using the entire annotated training set (5,828). Table II shows the results of these experiments. In both cases, we notice an increase in model performance as the amount of natural data introduced in the training process increases. Using the *Model-Iteration-4* approach consistently outperforms the model starting with the *Model-Iteration 0*, especially in the low-data regime. Pre-training with real images using pseudo-labels increases the performance of the segmentation model. In both cases, the F1-score increases by roughly 4% by adding just 25 annotated naturalistic samples, and by over 10% by adding 200 annotated samples. We train a model from scratch using a set of 200 labeled images and observe no valid predictions under the same training conditions. In contrast, our proposed approach can achieve high segmentation accuracy, bridging the performance gap with a model fine-tuned with the entire 5,828 images.

Figure 5 shows some examples of images where the seatbelt is segmented with reasonable accuracy. We observe the seatbelt is predicted with high confidence. There is minimum activation in areas where the seatbelt is not present. We observe the model works well with different views (e.g., Figs. 5(c) and 5(f)). It also works with passengers in the rear seat when the seatbelt is visible (Fig. 5(d)). Figure 6 shows some examples where our model fails. Figures 6(a), 6(b) and 6(c) show examples of false negatives. Here, the images have seatbelts that were not predicted by the model. We observe

TABLE II: Performance of the seatbelt segmentation approach after fine-tuning the model on annotated natural images. The pre-trained models include the iteration 0 (synthetic images), and iteration 4 (synthetic + real images with pseudo labels). LI represents labeled images.

LI	Model-Iteration 0				Model-Iteration 4			
	Prec.	Rec.	F1	IoU (%)	Prec.	Rec.	F1	IoU (%)
0	0.58	0.43	0.50	19.8	0.61	0.51	0.55	23.0
25	0.63	0.47	0.54	20.3	0.64	0.55	0.59	25.7
50	0.67	0.51	0.58	23.3	0.68	0.57	0.62	27.9
100	0.70	0.57	0.62	27.3	0.71	0.58	0.64	29.9
150	0.71	0.58	0.64	29.4	0.71	0.60	0.65	31.4
200	0.68	0.58	0.62	29.9	0.75	0.60	0.67	32.1
5,828	0.79	0.71	0.75	41.7	0.79	0.73	0.76	42.7

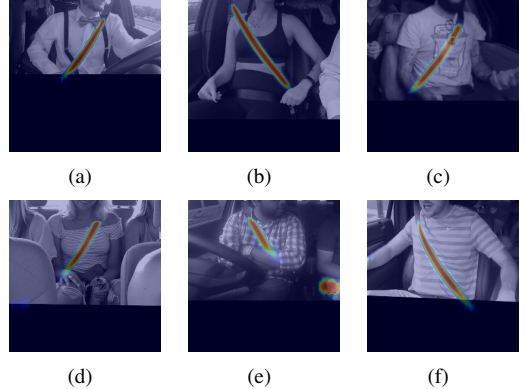


Fig. 5: Examples of successful prediction of seatbelts using our proposed approach.

that the model fails in these cases because of strong occlusions or reflections in the image. The contrast difference between the background and the seatbelt also plays a role. It is a more challenging problem when the color of the clothes of the driver or passengers is similar to the color of the seatbelt. Figures 6(d), 6(e) and 6(f) show images with false positive predictions. In these examples, the model incorrectly predicts a seatbelt where there is none. False positives appear in places where similar stripe-like patterns are present, such as the person’s arm (Fig. 6(d)) or steering wheel (Fig. 6(f)).

## VI. CONCLUSIONS

This paper presented a method to predict the seatbelt of a driver or passenger from an image that shows the upper body. This model can be used to predict if the driver is wearing the seatbelt, as well as to predict if the seatbelt is properly worn. We proposed a novel method to segment seatbelts that require minimal manual annotations. The model generates synthetic data resembling seatbelt patterns, which is used to train the base model. Naturalistic data is added in the training set with pseudo-labels to fine-tune the model, adapting the model to more efficiently work on real driving data. This pipeline of synthetic pre-training and iterative training with pseudo-labels requires no manual annotations or oversight. The addition of small amounts of annotated data for further fine-tuning the model increases the performance. The best performance is reached with a combination of synthetic pre-training, fine-

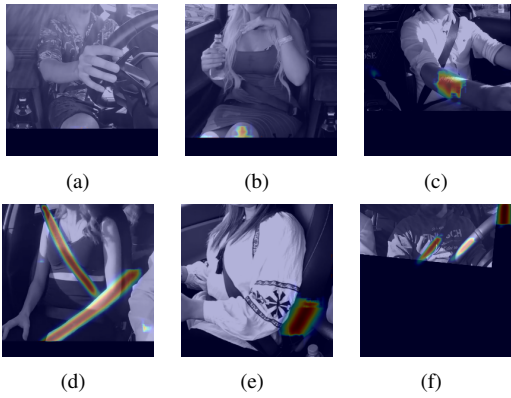


Fig. 6: Examples of incorrect seatbelt prediction. Figures (a)-(c) are false negatives, and Figures (d)-(f) are false positives.

tuning with pseudo-labels, and further training with labeled images. Adding even small amounts of labeled images (200 or less) increases the F1-score by up to 12%.

In the future, we will explore methods to merge an upper-body joint model and a seatbelt detection algorithm using multi-task learning approaches. Completing both tasks with a single network will provide computational flexibility to our system, making our solutions more appealing for in-vehicle applications. We observe some false predictions in the image due to similar patterns from other objects (e.g., the steering wheel). We can remove some false detections by adding contextual information about seat placement, ignoring predictions located in areas where seatbelts are not expected.

## REFERENCES

- [1] S. Jha and C. Busso, "Probabilistic estimation of the driver's gaze from head orientation and position," in *IEEE International Conference on Intelligent Transportation (ITSC)*, Yokohama, Japan, October 2017, pp. 1630–1635.
- [2] A. Rangesh, B. Zhang, and M. M. Trivedi, "Driver gaze estimation in the real world: Overcoming the eyeglass challenge," *arXiv preprint arXiv:2002.02077*, 2020.
- [3] S. Jha and C. Busso, "Probabilistic estimation of the gaze region of the driver using dense classification," in *IEEE International Conference on Intelligent Transportation (ITSC 2018)*, Maui, HI, USA, November 2018, pp. 697–702.
- [4] B. Vasli, S. Martin, and M. M. Trivedi, "On driver gaze estimation: Explorations and fusion of geometric and data driven approaches," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016, pp. 655–660.
- [5] S. Jha and C. Busso, "Estimation of driver's gaze region from head position and orientation using probabilistic confidence regions," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 59–72, January 2023.
- [6] M. Martin, S. Stuehmer, M. Voit, and R. Stiefelhagen, "Real time driver body pose estimation for novel assistance systems," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 1–7.
- [7] S. Y. Cheng and M. M. Trivedi, "Turn-intent analysis using body pose for intelligent driver assistance," *IEEE Pervasive Computing*, vol. 5, no. 4, pp. 28–37, 2006.
- [8] N. Li and C. Busso, "Analysis of facial features of drivers under cognitive and visual distractions," in *IEEE International Conference on Multimedia and Expo (ICME 2013)*, San Jose, CA, USA, July 2013, pp. 1–6.
- [9] —, "Predicting perceived visual and cognitive distractions of drivers with multimodal features," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 51–65, February 2015.
- [10] W. Li, Y. Cui, Y. Ma, X. Chen, G. Li, G. Zeng, G. Guo, and D. Cao, "A spontaneous driver emotion facial expression (defe) dataset for intelligent vehicles: Emotions triggered by video-audio clips in driving scenarios," *IEEE Transactions on Affective Computing*, 2021.
- [11] G. W. Trinca, "The influence of seatbelts use on road accident injury patterns," *Australian and New Zealand Journal of Surgery*, vol. 56, no. 1, pp. 13–17, January 1986.
- [12] O. J. Ogundele, A. Ifesanya, S. Adeyanju, and S. Ogunlade, "The impact of seat-belts in limiting the severity of injuries in patients presenting to a university hospital in the developing world," *Nigerian Medical Journal*, vol. 54, no. 1, pp. 17–21, January-February 2013.
- [13] M. Herath, P. Bautz, D. Parker, and C. Dobbins, "The importance of wearing a seatbelt correctly—a case report of blunt trauma causing complete shearing transection of the gastroduodenal junction," *International journal of surgery case reports*, vol. 72, pp. 197–201, 2020.
- [14] B. Zhou, L. Chen, J. Tian, and Z. Peng, "Learning-based seat belt detection in image using salient gradient," in *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2017, pp. 547–550.
- [15] H. Guo, H. Lin, S. Zhang, and S. Li, "Image-based seat belt detection," in *Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety*. IEEE, 2011, pp. 161–164.
- [16] B. Zhou, D. Chen, and X. Wang, "Seat belt detection using convolutional neural network bn-alexnet," in *International Conference on Intelligent Computing*. Springer, 2017, pp. 384–395.
- [17] A. Elihos, B. Alkan, B. Balci, and Y. Artan, "Comparison of image classification and object detection for passenger seat belt violation detection using NIR RGB surveillance camera images," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2018, pp. 1–6.
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, ser. Lecture Notes in Computer Science, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Munich, Germany: Springer Berlin Heidelberg, October 2017, vol. 9351, pp. 234–241.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NIPS 2012)*, vol. 25, Lake Tahoe, CA, USA, December 2012, pp. 1097–1105.
- [20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning (PMLR 2015)*, vol. 37, Lille, France, July 2015, pp. 448–456.
- [21] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.
- [22] S. Chun, N. Hamidi Ghalehjegh, J. Choi, C. Schwarz, J. Gaspar, D. McGehee, and S. Baek, "Nads-net: A nimble architecture for driver and seat belt detection via convolutional neural networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.
- [23] C. M. Ward, J. Harguess, and C. Hilton, "Ship classification from overhead imagery using synthetic data and domain adaptation," in *OCEANS 2018 MTS/IEEE Charleston*, 2018, pp. 1–5.
- [24] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Synthetic data and artificial neural networks for natural scene text recognition," *arXiv preprint arXiv:1406.2227*, 2014.
- [25] D.-H. Lee *et al.*, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, vol. 3, no. 2, 2013.
- [26] Y. Zou, Z. Zhang, H. Zhang, C.-L. Li, X. Bian, J.-B. Huang, and T. Pfister, "Pseudoseg: Designing pseudo labels for semantic segmentation," 2020. [Online]. Available: <https://arxiv.org/abs/2010.09713>
- [27] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba, and A. Oliva, "Places: An image database for deep scene understanding," *arXiv preprint arXiv:1610.02055*, 2016.
- [28] A. B. Jung *et al.*, "imgaug," <https://github.com/aleju/imgaug>, 2020, online; accessed 01-Feb-2020.
- [29] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: realtime multi-person 2d pose estimation using part affinity fields," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 1, pp. 172–186, 2019.