Technical Disclosure Commons

Defensive Publications Series

February 28, 2019

Training a machine-learning based object detector for use in photography

Ruiduo Yang

Marc Levoy

Training a machine-learning based object detector for use in photography

ABSTRACT

Cameras and other photography systems include the capability to detect objects of interest. Object detectors trained using sample data have a classification loss, e.g., due to insufficient training owing to a finite number of negative samples used during training. An increase in complexity of the object detector increases running time which necessitates a tradeoff between recall, precision, and speed. A common approach is to minimize the average loss across the entire training database. This disclosure proposes a new framework that takes the final image quality into account while training an object detector, by using a modified loss calculation function for the object detection framework used in photography. The framework enables better decisions regarding the various tradeoffs involved in loss calculation. The loss of classification during training is calculated by comparison of an image captured with and without successful detection of the object. The object detector, trained to take into account the impact of detecting or missing an object on a captured image, can improve the quality of captured images.

KEYWORDS

- Machine learning model
- Model training
- Loss function
- Object detection
- Face detection
- Group photo
- Camera

BACKGROUND

When capturing an image, cameras and other photography systems, such as those used in mobile devices, employ functionality to detect objects of interest, e.g., faces. The object detector functionality is often implemented using techniques such as Single Shot Detector (SSD) or Fast Region-based Convolutional Neural Networks (Fast R-CNN). All machine-learning based detectors exhibit a classification loss during training with manually supplied class labels. The loss is the confidence value produced by the last network layer, such as a fully connected layer followed by a softmax layer.

Ideally, a complex network ought to fit all the training data with no classification loss. However, this is not possible in practice since the number of negative samples could be infinite. Further, an increase in complexity increases running time, thus making a complex network unsuitable for use during image capture, e.g., by devices such as a mobile phones, which have limited processing capabilities. This necessitates a tradeoff between recall, precision, and speed. To this end, a common approach is to minimize the average loss across the entire training database.

Additionally, the importance of an individual sample could differ across situations. Existing approaches do not consider this difference and may not be optimal for photography use cases. For example, an object detector that fails to detect a backlit face can result in a bad photograph because of underexposure of the face whereas the end result may not be different in situations where the undetected face is in normal lighting. Similar issues affect other photography parameters, such as white balance, focus, and noise.

DESCRIPTION

This disclosure proposes a new framework that takes the final image quality into account while training an object detector, by using a modified loss calculation function for the object detection framework used in photography. The modified loss calculation enables making better decisions regarding the various tradeoffs involved in loss calculation performed for object detection in photography. The classification loss during training is calculated by considering a penalty that is based on comparing the final image with and without successful detection of the object.

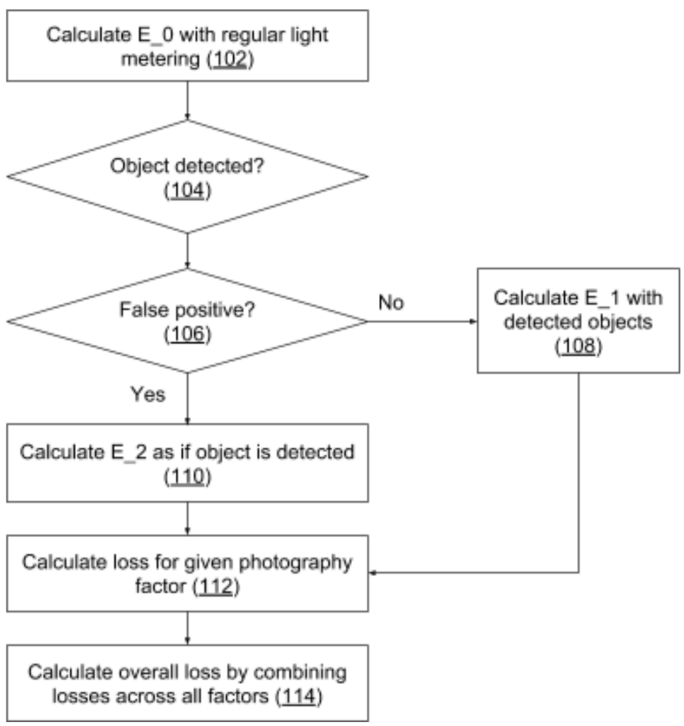

Fig. 1: Example process to calculate loss

Fig. 1 shows an example process, per techniques described in this disclosure. Photography factor values (E_0) are calculated (102) using the regular light metering of the camera as if no objects are detected. For example, the factors can include exposure, white balance, focus, noise, etc. The calculation is performed using a camera metering algorithm, such as center weighted metering. The calculation can be done with the image in camera raw space, linear space, or non-linear space, depending on availability of training data. This calculation can be based on many different 3A algorithms.

If objects are detected (104), the data is checked for a false positive (106). If there is no false positive, the photography factor values (E_1) are calculated using an object-based algorithm applied to the detected objects (108). When a false positive occurs, photography factor values (E_2) are calculated as if an object is detected (110). For example, a false positive occurs when a negative sample detected as including an object.

The calculated values - E_0 and E_1 or E_0 and E_2 - are used to calculate a final classification loss for the given photography factor via a suitable function (112). The final classification loss is calculated via a suitable function F(f(E_0, E_1), L) or F(f(E_0, E_2), L), respectively, where L is a function to calculate a general classification loss and f is the function to calculate the appropriate weight for L. The function F generates the final loss that indicates how the given sample affects an image. The individual factor classification losses calculated for each of the photography factor values are combined to yield the overall classification loss for the object detector (114).

Applying the approach described in this disclosure results in notably different ratios of detection rate/false positive in various photography scenes, e.g., when capturing images of a group, compared to the differences in the corresponding ratios when employing conventionally trained object detector systems. The loss calculations can use general loss such as perceptual loss that measures the visual difference between two images. A key aspect of the proposed techniques is combining the loss with the results of object detection such that the loss calculations are based on the difference between images with or without the object being detected. Such consideration of the impact of detecting or missing an object on the final image results in notable improvements to image quality of images captured using a camera that implements such techniques.

One alternative is to use an object detector with greater complexity to fit more training data. However, even with increased complexity and running time, the more difficult cases may not be solved. In contrast, the approach proposed in this disclosure provide improvements in image quality without an increase in running speed and model complexity, and can achieve improvements in loss calculations for object detection functionality of camera systems.

The techniques can be implemented in any training context, e.g., server-based training, or training on a client device. A machine-learning model trained with use of the loss function as described can be employed for object detection in any image capture device, e.g., smartphone, wearable device, digital camera, and other devices that include object detection during image capture.

CONCLUSION

Cameras and other photography systems include the capability to detect objects of interest. Object detectors trained using sample data have a classification loss, e.g., due to insufficient training owing to a finite number of negative samples used during training. An increase in complexity of the object detector increases running time which necessitates a tradeoff between recall, precision, and speed. A common approach is to minimize the average loss across the entire training database. This disclosure proposes a new framework that takes the final image quality into account while training an object detector, by using a modified loss calculation function for the object detection framework used in photography. The framework enables better decisions regarding the various tradeoffs involved in loss calculation. The loss of classification during training is calculated by comparison of an image captured with and without successful detection of the object. The object detector, trained to take into account the impact of detecting or missing an object on a captured image, can improve the quality of captured images.