

ion. The positive samples correspond to predictions from alt boxes matched with any of ground truth traffic sign, 'matched' means that the intersection over union (IoU) between the default box and the ground truth target is larger 0.5. If a default box is not matched with any ground truth set, then the prediction from the default box is regarded as negative sample.

We use the softmax loss as shape classification loss  $L_{shape}(s)$  for all positive and negative samples, and use both  $\ell_1$  loss [40] as vertex regression loss  $L_{vertex}(\Delta p^i)$  for positive samples. For negative samples, the vertex regression loss is not assigned. The overall training loss for single image is the sum of the shape classification loss  $L_{shape}(s)$  and the vertex regression loss  $L_{vertex}(\Delta p)$  as follows,

$$L_{overall}(\mathbf{s}, \Delta \mathbf{p}^i) = \frac{1}{K_p + K_n} \lambda_s \sum_{K_p + K_n}^{K_p + K_n} L_{shape}(\mathbf{s}) + \frac{1}{K_p} \lambda_v \sum_{K_p}^{K_p} \sum_{i=1}^4 L_{vertex}(\Delta \mathbf{p}^i), \quad (4)$$

where  $K_p$  and  $K_n$  are the number of positive and negative samples respectively, and  $\lambda_s$  and  $\lambda_v$  are loss weights for the shape classification loss and the vertex regression loss respectively. For notational simplicity, we omit the index of a default box inside each loss function.

Since multiple default boxes of various aspect ratios are assigned to each unit of classification and regression layers, it is usual that the total number of default boxes is significantly larger than the number of positive samples, and most of the default boxes become negative samples. To resolve this severe imbalance between positive and negative samples, we employ *hard negative mining* technique [6], [41] to select only few of

ive and negative samples are used for training. The positive samples correspond to predictions from default boxes matched with any of ground truth traffic sign. 'matched' means that the intersection over union (IoU) between the default box and the ground truth target is larger than 0.5. If a default box is not matched with any ground truth target, then the prediction from the default box is regarded as a negative sample.

We use the softmax loss as shape classification loss for all positive and negative samples, and use both  $\ell_1$  loss [40] as vertex regression loss  $L_{vertex}(\Delta \mathbf{p}^i)$  for positive samples. For negative samples, the vertex regression loss is not assigned. The overall training loss for single image is the sum of the shape classification loss  $L_{shape}(\mathbf{s})$  and the vertex regression loss  $L_{vertex}(\Delta \mathbf{p})$  as follows,

$$L_{\text{overall}}(\mathbf{s}, \Delta \mathbf{p}^i) = \frac{1}{K_p + K_n} \lambda_s \sum_{K_p+K_n}^{K_p+K_n} L_{\text{shape}}(\mathbf{s}) + \frac{1}{K_p} \lambda_v \sum_{K_p}^{K_p} \sum_{i=1}^4 L_{\text{vertex}}(\Delta \mathbf{p}^i), \quad (4)$$

where  $K_p$  and  $K_n$  are the number of positive and negative samples respectively, and  $\lambda_s$  and  $\lambda_v$  are loss weights for the shape classification loss and the vertex regression loss respectively. For notational simplicity, we omit the index of a default box inside each loss function.

Since multiple default boxes of various aspect ratios are assigned to each unit of classification and regression layers, it is usual that the total number of default boxes is significantly larger than the number of positive samples, and most of the default boxes become negative samples. To resolve this severe imbalance between positive and negative samples, we employ *hard negative mining* technique [61, 44] to select only few of

ation. The positive samples correspond to predictions from  
ult boxes matched with any of ground truth traffic sign.  
re 'matched' means that the intersection over union (IoU)  
ween the default box and the ground truth target is larger  
0.5. If a default box is not matched with any ground truth  
et, then the prediction from the default box is regarded as  
egative sample.

We use the softmax loss as shape classification loss  $L_{\text{shape}}(\mathbf{s})$  for all positive and negative samples, and use both the loss [40] as vertex regression loss  $L_{\text{vertex}}(\Delta \mathbf{p}^i)$  for positive samples. For negative samples, the vertex regression loss is not assigned. The overall training loss for single image is the sum of the shape classification loss  $L_{\text{shape}}(\mathbf{s})$  and the vertex regression loss  $L_{\text{vertex}}(\Delta \mathbf{p})$  as follows,

$$L_{overall}(\mathbf{s}, \Delta \mathbf{p}^i) = \frac{1}{K_p + K_n} \lambda_s \sum_{K_p + K_n}^{K_p + K_n} L_{shape}(\mathbf{s}) + \frac{1}{K_p} \lambda_v \sum_{i=1}^{K_p} \sum_{v=1}^4 L_{vertex}(\Delta \mathbf{p}^i), \quad (4)$$

where  $K_p$  and  $K_n$  are the number of positive and negative samples respectively, and  $\lambda_s$  and  $\lambda_v$  are loss weights for the shape classification loss and the vertex regression loss respectively. For notational simplicity, we omit the index of a default box inside each loss function.

Since multiple default boxes of various aspect ratios are assigned to each unit of classification and regression layers, it is usual that the total number of default boxes is significantly larger than the number of positive samples, and most of the default boxes become negative samples. To resolve this severe imbalance between positive and negative samples, we employ *hard negative mining* technique [61, 441] to select only few of

boxes matched with any of ground truth target boxes. "matched" means that the intersection over union of the default box and the ground truth target is greater than 0.5. If a default box is not matched with any ground truth target, then the prediction from the default box is regarded as a false sample.

se the softmax loss as shape classification  
 a) for all positive and negative samples, a  
 1) loss [40] as vertex regression loss  $L_{\text{vertex}}$   
 positive samples. For negative samples, the  
 on loss is not assigned. The overall training  
 e image is the sum of the shape classification  
 s) and the vertex regression loss  $L_{\text{vertex}}(\Delta \hat{v}_i$

$$loss(s, \Delta p^i) = \frac{1}{K_p + K_n} \lambda_s \sum_{K_p + K_n} L_{shape}(s) + \frac{1}{K_p} \lambda_v \sum_{K_p} \sum_{i=1}^4 L_{vertex}(\Delta p^i).$$

$K_p$  and  $K_n$  are the number of positive and negative examples respectively, and  $\lambda_s$  and  $\lambda_v$  are loss weights for shape classification loss and the vertex regression loss respectively. For notational simplicity, we omit the index of the box inside each loss function.

Since multiple default boxes of various aspect ratios are tied to each unit of classification and regression layers, it is usual that the total number of default boxes is significant or more than the number of positive samples, and most of the default boxes become negative samples. To resolve this severe

tion. The positive samples correspond to predictions from alt boxes matched with any of ground truth traffic sign, i.e. 'matched' means that the intersection over union (IoU) between the default box and the ground truth target is larger than 0.5. If a default box is not matched with any ground truth target, then the prediction from the default box is regarded as a negative sample.

We use the softmax loss as shape classification loss  $L_{\text{shape}}(s)$  for all positive and negative samples, and use both  $L_1$  loss [40] as vertex regression loss  $L_{\text{vertex}}(\Delta p^i)$  for positive samples. For negative samples, the vertex regression loss is not assigned. The overall training loss for single image is the sum of the shape classification loss  $L_{\text{shape}}(s)$  and the vertex regression loss  $L_{\text{vertex}}(\Delta p)$  as follows,

$$L_{overall}(s, \Delta P^i) = \frac{1}{K_p + K_n} \lambda_s \sum_{K_p + K_n} L_{shape}(s) + \frac{1}{K_p} \lambda_v \sum_{K_p} \sum_{i=1}^4 L_{vertex}(\Delta P^i), \quad (4)$$

where  $K_p$  and  $K_n$  are the number of positive and negative samples respectively, and  $\lambda_s$  and  $\lambda_v$  are loss weights for the shape classification loss and the vertex regression loss respectively. For notational simplicity, we omit the index of a default box inside each loss function.

Since multiple default boxes of various aspect ratios are assigned to each unit of classification and regression layers, it is usual that the total number of default boxes is significantly larger than the number of positive samples, and most of the default boxes become negative samples. To resolve this severe imbalance between positive and negative samples, we employ hard negative mining technique [61, 44] to select only few of