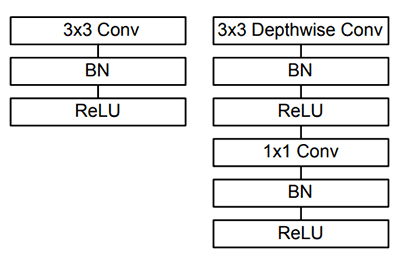
**SSD**

* For example, in Conv4\_3, we apply four 3 × 3 filters to map 512 input channels to 25 output channels (21 scores for each class plus one boundary box).

**Имплементација**

* MobileNets differ from traditional CNNs through the usage of depthwise separable convolution. The general idea behind depthwise separable convolution is to split convolution into two stages which allows us to actually reduce the number of parameters in our network:

A 3×3 depthwise convolution.

Followed by a 1×1 pointwise convolution.

(Left) Standard convolutional layer with batch normalization and ReLU. (Right) Depthwise separable convolution with depthwise and pointwise layers followed by batch normalization and ReLU

* The model weights file and a protobuf text file contains the model configuration. The weight file has a .pb extension which is a protobuf file containing all the pre-trained weights. The model configuration is held in the protobuf text file, which has a .pbtxt file extension.
* Za Caffe: The .prototxt file(s) which define the model architecture (i.e., the layers themselves). The .caffemodel file which contains the weights for the actual layers
* Mean subtraction is used to help combat illumination changes in the input images in our dataset.
* Прерачунате су две променљиве: једна у којој је смештена жељена х координата средишње тачке између очију на коначној слици (половина жељене ширине слике), и друга у којој се налази у координата те тачке. Транслациона компонента матрице је ажурирана тако што су од ове две вредности одузете одговарајуће координате тренутне централне тачке.
* Erozija: it erodes away the boundaries of foreground object (Always try to keep foreground in white)
* For every pixel, the same threshold value is applied. If the pixel value is smaller than the threshold, it is set to 0, otherwise it is set to a maximum value. The function cv.threshold is used to apply the thresholding. The first argument is the source image, which should be a grayscale image. The second argument is the threshold value which is used to classify the pixel values. The third argument is the maximum value which is assigned to pixel values exceeding the threshold. OpenCV provides different types of thresholding which is given by the fourth parameter of the function. Basic thresholding as described above is done by using the type cv.THRESH\_BINARY.
* Erode the resulting image to remove backlights
* ResNet (2015) Veoma duboka mreža koju čini 152 sloja, gde je predložen radni okvir (eng. framework) učenja dubokih ostataka (eng. deep residual learning) kojim olakšava optimizaciju i konvergenciju ekstremno dubokih mreža. Naziv ResNet, arhitektura je dobila od skraćenice za „Residual Network”.



* Distorzija kamere: The presence of the radial distortion manifests in form of the “barrel” or “fish-eye” effect. Tangential distortion occurs because the image taking lenses are not perfectly parallel to the imaging plane.

Да би се дефинисала mAP мера, треба прво дефинисати прецизност (енг. precision) и комплетност (енг. recall). Прецизност представља проценат позитивно класификованих података који су заиста позитивни (), док комплетност представља проценат исправно класификованих позитивних података у односу на укупан број позитивних података (). TP (True Positive) су позитивни подаци који су исправно детектовани као позитивни, FP (False Positive) су негативни подаци који су погрешно детективани као позитивни и FN (False Negative) су позитивни подаци који су погрешно детектовани као негативни. Још једна значајна мера је IOU (Intersection Over Union). IOU представља однос између пресека и уније два правоугаоника на слици (предиктованог правоугаоника и ручно означеног правоугаоника). Предикција се означава као TP ако даје IOU већи од неког прага (рецимо 0.5). Average Precision (AP) се дефинише као површина испод графика зависности прецизности од комплетности. In Pascal VOC2008, an average for the 11-point AP is calculated. For later Pascal VOC competitions, VOC2010–2012 samples the curve at all unique recall values (r₁, r₂, …), whenever the maximum precision value drops. mAP (mean average precision) is the average of AP. In some context, we compute the AP for each class and average them. But in some context, they mean the same thing. For example, under the COCO context, there is no difference between AP and mAP.

True Positives (TP): The model predicted a label and matches correctly as per ground truth.

True Negatives (TN): The model does not predict the label and is not a part of the ground truth.

False Positives (FP): The model predicted a label, but it is not a part of the ground truth (Type I Error).

False Negatives (FN): The model does not predict a label, but it is part of the ground truth. (Type II Error).