Senior Design Project 1

Madhana Gopalapuram Ramesh Rathish Aadiraja
302457

Segmentation and recognition of information printed on identification cards

Different methods for localization of ID card from the input image:

[1] A method based on detection of quadrilateral of document border in the image. It is a combination of contour and region based approach. It is a modified contour approach in which the contours detected are ranked according to contrast between areas inside and outside the border. The quadrilateral with the highest score is selected and it is compared with the ground truth quadrilateral based on Jaccard Index (intersection over union). If the Jaccard index value is greater than the threshold coefficient then the quadrilateral is considered as a correct one, else it is not considered.

[2] A method which simultaneously locates the document and recognizes its class then, in next steps the document nature, country, version and the visible side is determined. For every ID card type one reference model is created and the keypoints are extracted and classified by SURF method. Each reference model is indexed with random KD trees. From the query image the keypoints are extracted and matched against all reference models at the same time with a matching score. Then the reference models are reverse matched against the query image and compared with the symmetric mapping of couple of keypoints of direct and reverse matching, histogram of orientation difference between these points and geometric transformation using RANSAC and given a score for each then compared with the previous score. The one with the highest score is considered to be a valid reference model. Then the quadrilateral of the document is detected and verified with set of validations. This approach is the baseline for the approach in the paper [8].

[3] A method using a sliding window to detect every region of the image if an object of interest is located or not. For each window the occurrences of Gradient Orientation(HOG) in the certain portion of the image is calculated. Then each window is classified by SVM if it contains the document or not.

[3] A method based on DenseNet10. The architecture includes Bach normalization, ReLU, Convolutional layer and dropout. It is implemented by using 3 blocks and 10 layers making it more lightweight for mobile device applications. This is done by having one up-sampling path which has one transition up(1 TU) and one down-sampling path with one transition down(1 TD) .

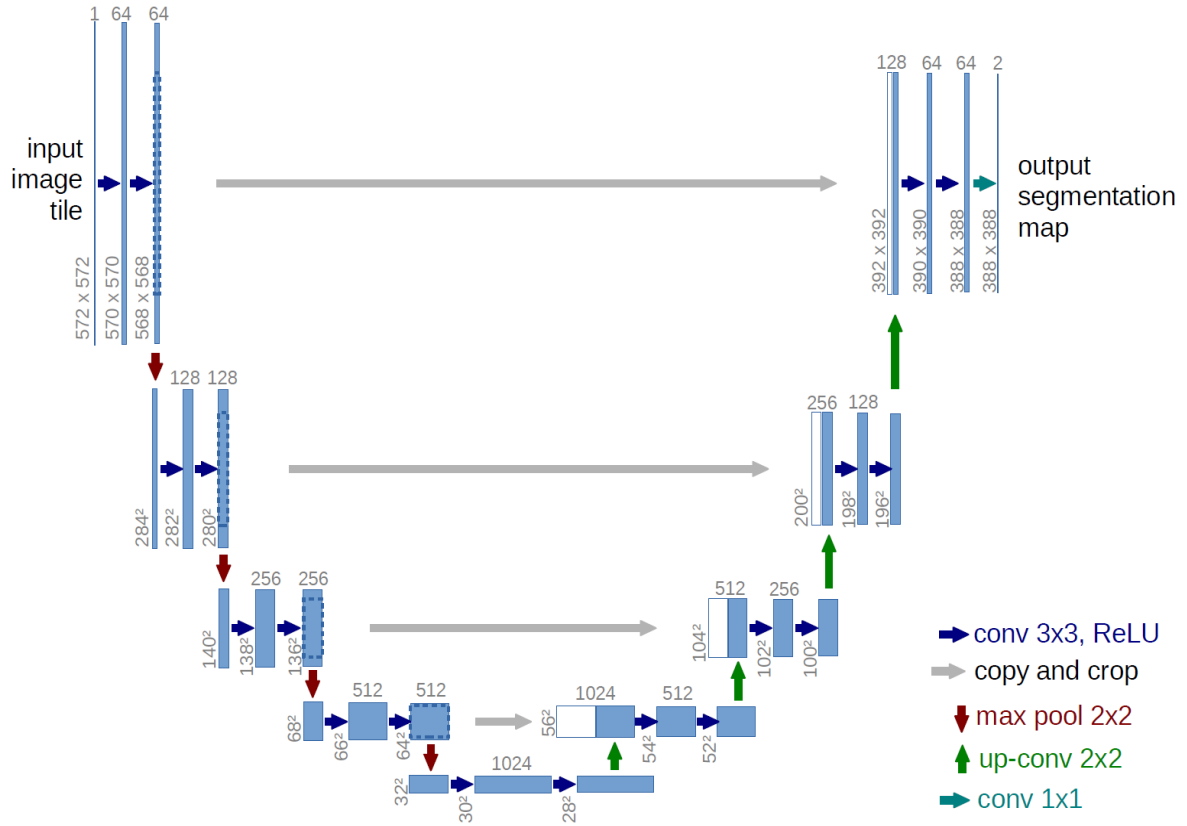
[4] A method which uses the information of colour difference of the ID card and the background for localization of the ID card. It works by adjusting the vertices of the documents iteratively using data of pixels sampled in outer region of the ID card in the image. It has a priori assumption that the document is approximately in the centre of the

input image. Another set of vertices start at 30% of the edges of the document which occupies 70% of the whole images and adjusts iteratively until it finds the perfect matching vertices that contains the ID card in the entire image. After the localization it is classified using CNN.

[5] A method that spots the ID card and accurately localizes it from the original image using specific ID document features. It requires a classification a priori along with the list of predefined models. Classification is performed by local key descriptor matching process (SURF) algorithm and RANSAC is used after this step to estimate transformation matrix. A multi-hypothesis approach which runs different crop solutions in parallel and selects the one with the highest score as the cropped document. The main features that are used in the feature extraction and cropping methods are keypoints in the image, vanishing points, MRZ, document border, document corners, photo of a person's face in the document, landmarks, headpose of the person in the document, logo in the document, variable and invariable fields of the document. Among these features a best set of hypothesis of features are selected to produce the best crop of ID document in any different situation.

Localization using UNet:

[6] UNet is a convolutional neural network architecture that was primarily developed for biomedical image segmentation by Olaf Ronneberger, Philipp Fischer, and Thomas Brox. It can successfully perform semantic segmentation with just scant amount of data. This architecture can be viewed in two main parts, the encoder and decoder path. The encoder path is the contracting path which captures the context and the decoder path is the expanding path which enables precise localization.



At first the input image is taken with the dimensions of 572 x 572 x 1 which is the length, width and number of channels of the image. Then a 3 x 3 convolution operation is applied to this image with 64 filters, no padding and with stride of 1. This is followed by a non-linear activation function, rectified linear unit (ReLU). Again for this feature map the same 3 x 3 convolution operation is applied with a ReLU function. After these two convolution operations a 2 x 2 maxpooling operation with stride of 2 is performed, which reduces the length and width of the feature map by half. Now these three steps are repeated for three times only with a small change where the number of filters doubles in each downsampling step. After these steps the resulting feature map is of dimension 32 x 32 x 512 which is followed by two 3 x 3 convolution operations, ReLU and with double the number of filters used in the previous step, resulting with dimension 28 x 28 x 1024. Now, instead of maxpooling, a 2 x 2 up convolutions operation is performed which halves the number of feature channels in every step so the dimensions are changed to 54 x 54 x 512 . This resultant feature map is concatenated with the result that was obtained in each step of downsampling, before the maxpooling operation, resulting in a feature map of dimension 54 x 54 x 1024. Again two 3 x 3 convolution operations is performed with ReLU followed by a 2 x 2 up convolution and concatenated with the corresponding result from the downsampling step. Finally from the feature map of dimension 388 x 388 x 64 a 1 x 1 convolution operation is performed to get the final desired number of classes with the dimension of 388 x 388 x 2. The output image has two channels in which one channel is for the foreground class and the other for the background class. Due to the unpadded convolution operations the final output image is smaller than the input image. This architecture consists a total of 23 convolution layers. It is necessary that the length and width of the image should be an even number and of equal size before maxpooling operations.

This approach can also be applied for sematic segmentation of ID card from the original image, it is also used in [9]. This method classifies each and every pixel of the input image into two categories, the one that contains ID card and the one that does not contain the ID card (background). To train this model, the images in the dataset can be masked for the area under the quadrilateral formed by the ground truth vertices or can be annotated manually. Since the datasets that are available for images of ID cards are less comparatively to other datasets due to privacy, this architecture can be trained with the availability of less number of images with precise segmentation results.

[7] The accuracy of this localization can be verified by finding the value of a Generalized Intersection over Union which is Area of overlap divided by Area of union of the ground truth mask and the predicted mask by the UNet algorithm.

.

Bibliography:

[1] Daniil V. Tropin, Sergey A. Ilyuhin, Dmitry P. Nikolaev and Vladimir V. Arlazarov. Approach for document detection by contours and contrasts. https://arxiv.org/pdf/2008.02615.pdf

[2] Ahmad-Montaser Awal, Nabil Ghanmi. Complex Document Classification and Localization Application on Identity Document Images. https://hal.inria.fr/hal-01660504/document

[3] Rodrigo Lara, Andres Valenzuela, Daniel Schulz, Juan Tapia and Christoph Busch. Towards an Efficient Semantic Segmentation Method of ID Cards for Verification Systems. https://arxiv.org/pdf/2111.12764.pdf

[4] Filippo Attivissimo, Nicola Giaquinto, Marco Scarpetta and Maurizio Spadavecchia. An Automatic Reader of Identity Documents. https://arxiv.org/ftp/arxiv/papers/2006/2006.14853.pdf

[5] Guillaume Chiron, Nabil Ghanmi, and Ahmad Montaser Awal. ID documents matching and localization with multi-hypothesis constraints. https://hal.archives-ouvertes.fr/hal-03219532/document

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://arxiv.org/pdf/1505.04597.pdf

[7] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, Silvio Savarese. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. https://arxiv.org/pdf/1902.09630.pdf

[8] Natalya Skoryukina, Vladimir V. Arlazarov, and Dmitry P. Nikolaev. Fast method of ID documents location and type identification for mobile and server application. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8978136&tag=1

[9] Alejandra Castelblanco, Jesus Solano, Christian Lopez, Esteban Rivera, Lizzy Tengana, Martín Ochoa. Machine Learning Techniques for Identity Document Verification in Uncontrolled Environments: A Case Study. https://link.springer.com/content/pdf/10.1007%2F978-3-030-49076-8_26.pdf