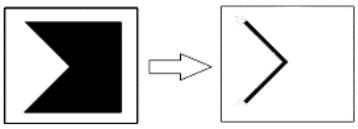
Video Signals

Date: 24 June 2026

Ex.1.[9 pts] Given the following image where the white pixels have a value of 1 and the black pixels have a value of 0 propose an edge extractor filter and describe the whole algorithm in order to get the resulting image on the right where the black line has pixels with value 0 and the background pixels have value 1. Furthermore, describe the effects on the corners and how to handle them.



Ex.2. [11 pts] Given the following black and white image (where black pixels have a value of 1 and white pixels have a value of 0), we need to apply the morphological skeletonization in order to describe its structure.
The structuring element is a 3x3 cross on the right.

0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	1	1	1	0	0	0
0	0	0	1	1	1	1	1	0	0	0
0	0	0	1	1	1	1	1	1	1	0
0	0	0	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	1	1	0
0	0	0	1	1	1	1	1	1	1	0
0	0	0	1	1	1	1	1	1	1	0
0	0	0	1	1	1	1	1	0	0	0
0	0	0	1	1	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0

0	1	0
1	1	1
0	1	0

[4 pts] Describe the procedure and the steps in order to get the skeleton.
[7 pts] Calculate the skeleton and describe the procedure, starting from it, to rebuild the original image.

Ex. 3 is overleaf

Es.3. [12 pts to be solved writing MATLAB code on the paper]

When a paper document is acquired using a scanner or a camera, it may appear slightly rotated with respect to the horizontal direction. This unwanted rotation is called skew and can reduce readability or negatively affect further document analysis tasks.

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

In this exercise, you will implement a simple automatic skew correction method based on horizontal projection profiles. The idea is that, when text lines are correctly aligned, the sum of text pixels along each image row produces strong peaks. Therefore, the best correction angle can be estimated by finding the rotation that maximizes the variance of the horizontal projection profile.
Starting from the input image paper.png, write a MATLAB script able to perform the following steps:

a) Read the image paper.png, convert it to double representation, and display it.
b) Convert the image to grayscale.
c) Apply a Gaussian smoothing filter with kernel size 5x5 and standard deviation equal to 1.
d) Create a binary mask of the text so the text pixels will appear as white and the background as black.
e) Define a vector of candidate angles from -20° to +20°, with a step of 0.5° and initialize a score vector with the same size.
f) For each candidate angle:
   I. Rotate the binary mask using nearest-neighbor interpolation.
   II. Compute the horizontal projection profile by summing the binary image row by row.
   III. Compute a score as the variance of the projection profile and save it into the score vector.
g) Estimate the best correction angle as the one that maximizes the score.
h) Rotate the original image using the estimated angle and bicubic interpolation. Display the corrected image.

List of possible Matlab functions

```
figure
im2double
imread
rgb2gray
imcrop
imfilter
imrotate
imopen
rgb2hvs
histeq
hist
imshow
fspecial
imerase
strel
hsv2rgb
imbinarize
```

Solutions

Es.1

A possible solution could be: starting from an image like this one:

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	0	0
0	0	0	1	1	1	1	1	1	1	1	1	1	0	0
0	0	0	0	1	1	1	1	1	1	1	1	1	0	0
0	0	0	0	0	1	1	1	1	1	1	1	1	0	0
0	0	0	0	0	0	1	1	1	1	1	1	1	0	0
0	0	0	0	0	0	0	1	1	1	1	1	1	0	0
0	0	0	0	0	0	1	1	1	1	1	1	1	0	0
0	0	0	0	0	1	1	1	1	1	1	1	1	0	0
0	0	0	0	1	1	1	1	1	1	1	1	1	0	0
0	0	0	1	1	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

We can apply a diagonal North-East Sobel filter:

0	-1	-2
1	0	-1
2	1	0

This produces the following filtered image:

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	3	3	3	3	3	3	3	3	3	3	2	0
0	-1	-1	1	3	3	3	3	3	3	3	3	4	3	0
0	-2	-4	-4	-2	0	0	0	0	0	0	0	3	3	0
0	0	-2	-4	-4	-2	0	0	0	0	0	0	3	3	0
0	0	0	-2	-4	-4	-2	0	0	0	0	0	3	3	0
0	0	0	0	-2	-4	-4	-2	0	0	0	0	3	3	0
0	0	0	0	0	-2	-3	-1	0	0	0	0	3	3	0
0	0	0	0	0	0	0	0	0	0	0	0	3	3	0
0	0	0	0	0	0	0	0	0	0	0	0	3	3	0
0	0	0	0	0	0	0	0	0	0	0	0	3	3	0
0	0	0	0	0	0	0	0	0	0	0	0	3	3	0
0	-1	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	0	1	0
0	-2	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

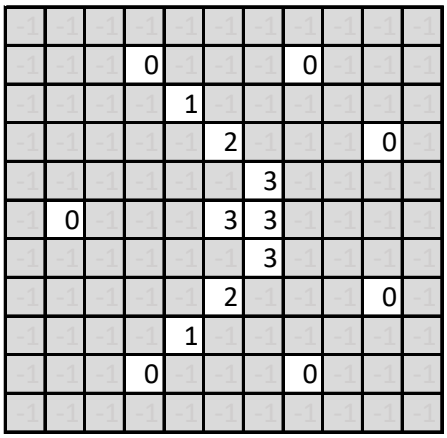
We can then apply the vertically flipped version of the filter, obtaining a vertically symmetric result.

Finally, we combine (take the union of) the two results by applying a threshold and keeping only values smaller than -4.

With this approach, some pixels at the corners are lost, so special attention should be paid to handling those cases.

Es.2

For the skeleton procedure, check the lectures recordings and slides.
The result will be:



The reconstruction will be union of the results of dilating the skeleton elements with the structuring element as many times as indicated in the skeleton and adding, as the final steps, the ‘leaves’, i.e., the elements indicated by ‘0’.

Es.3

```
clc
close all
clear all

% a) Read the image
I = im2double(imread('paper.png'));

figure
imshow(I)
title('Original image')

% b) Convert to grayscale
I_gray = rgb2gray(I);

% c) Gaussian smoothing
G = fspecial('gaussian', [5 5], 1);
I_filt = imfilter(I_gray, G, 'conv', 'same', 'symmetric');

% d) Text binarization
% The text is dark, so pixels below the threshold are selected
BW = ~imbinarize(I_filt);

% e) Search for the best angle using projection profiles
angles = -20:0.5:20;
scores = zeros(size(angles));

%f)
for k = 1:length(angles)

    theta = angles(k);

    % f.I Rotate the binary mask
    BW_rot = imrotate(BW, theta, 'nearest', 'crop');
```

```
    % f.II Horizontal projection profile
    projection = sum(BW_rot, 2);

    % f.III Score: when text lines are horizontal,
    % the projection profile has sharper peaks
    scores(k) = var(projection);

end

% g) Estimate the angle
[~, idx] = max(scores);
best_angle = angles(idx);

% h) Correct the original image
I_corrected = imrotate(I, best_angle, 'bicubic');

figure
imshow(I_corrected)
```