#### **DIEF- Dipartimento di Ingegneria Enzo Ferrari**

# **Handwritten-notes Denoising**

#### **Authors**

Olmo Baldoni, Cristian Bellucci, Danilo Caputo

325524@studenti.unimore.it, 322906@studenti.unimore.it, 246019@studenti.unimore.it



## Introduction

#### • Challenge:

• Handwritten notes, especially on lined or squared sheets, are intrisacally noisy for automatic extraction of text and equations.

#### Solution:

Train a Convolutional Neural Network (CNN) to remove square grids from handwritten notes, resulting
in a text-only, noise-free image.

#### Methodology:

- Utilize UNet
- Create synthetic dataset for the training process

#### Additional Tools:

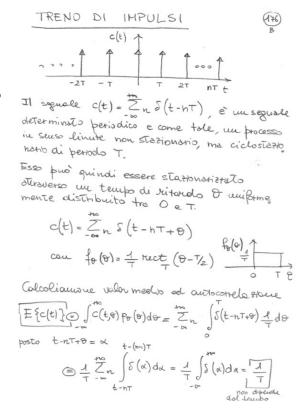
- An approach to crop and warp images of hand-taken notes from photos.
- A retrieval system for automating the collection of images to be processed by the denoising network.



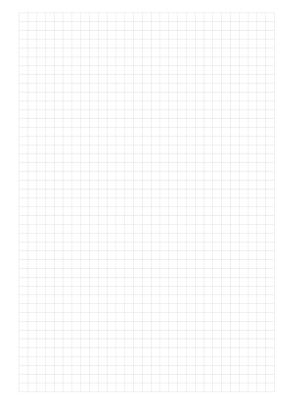
# Synthetic Dataset

- 132000 synthetic images obtained by a Data Augmentation process.
- 3300 images of handwritten notes on blank sheets
- 70 different template of different grids used as background
  - 40 of each randomly applied
- The grids G(x, y) were overlapped to the notes images N(x, y) using the max operator over the negative gray scale images:

$$I_{neg}(x,y) = \max(G_{neg}(x,y), N_{neg}(x,y))$$
$$I(x,y) = 255 - I_{neg}(x,y)$$



(a) Example of handwritten note



(b) Example of template



# Data Augmentation

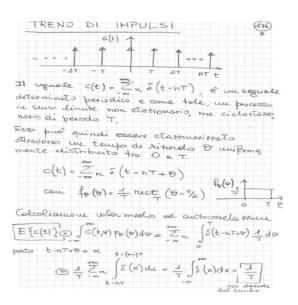
 Making synthetic images more realistic

#### Applied:

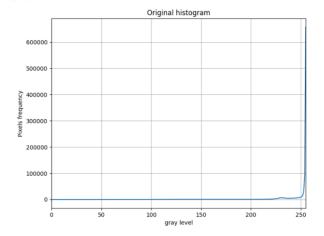
 Sinusoidal distortions that make the pixels of the templates slightly wavy

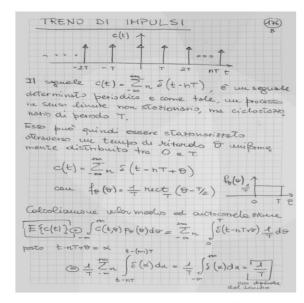
$$x_{new} = x + A \sin \frac{y}{k}$$
$$y_{new} = y + A \sin \frac{x}{k}$$

 Elliptical masks randomly applied to one or more regions of the image to alter the brightness

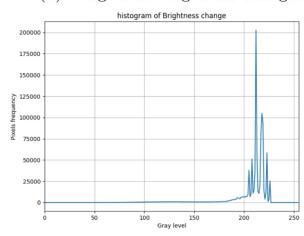


(a) Image without brightness changes





(b) Image with brightness changes





# Image cropping and warping

• Crop a sheet of paper in a image and warping it to original resolution.

#### First step:

- detecting sheet contours
  - morphological operation combined with Canny algorithm.

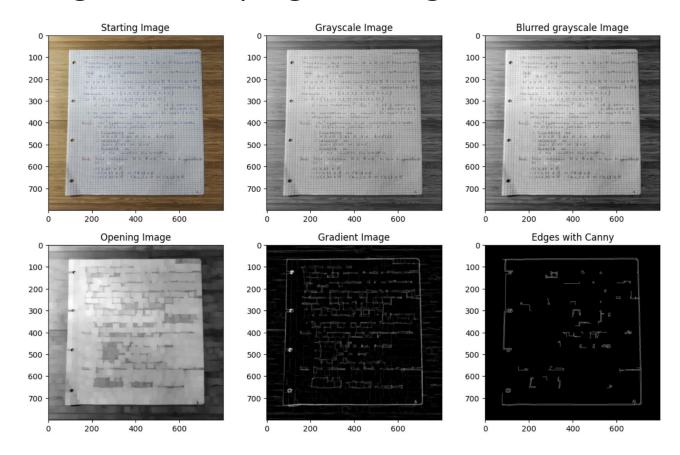


Figure 4: Pipeline of edge detection

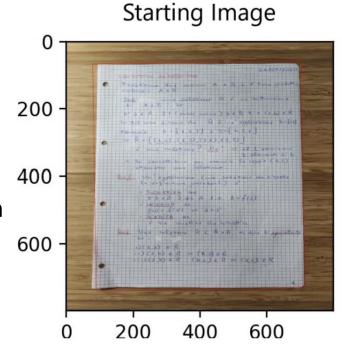


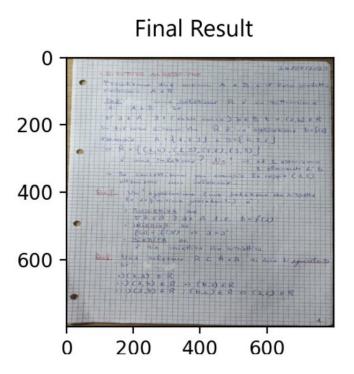
# Image cropping and warping

• Crop a sheet of paper in a image and warping it to original resolution.

#### **Second step:**

- Perspective Transformation
  - Extraction of corners coordinates of contour with maximum area
  - Map the detected sheet to corner of original image

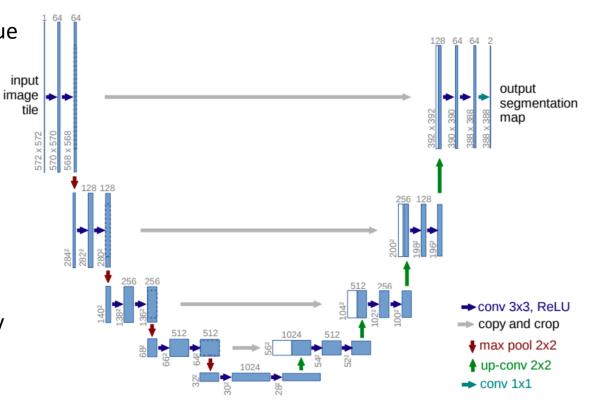






#### **Choosing UNet for Grid Removal**

- Previous methods like Fourier Transform were impractical due to:
  - Manual parameter tuning
  - Lack of automation
  - time consumption
- The need for a scalable and generalized solution.
- Switched to a deep learning approach using UNet for grid removal.
- UNet chosen due to its effectiveness in segmentation, ability to downsample and upsample.
- Suitable for complex images, and has the ability to maintain clarity of text and fine details.





#### **Train&Test set Composition**

- Image resolutions tested: 256, 512, and 1024 pixels
- Dataset divided into 2/3 train and 1/3 test split
- Overfitting issue:
  - Model primarily learning to remove synthetic grids
- Addressing overfitting:
  - Increased number of grids
  - data augmentation
  - expanded base images and grid variety



#### **Training Approach and Optimizers**

- Distributed training to handle large images efficiently.
- Use of Mean Squared Error (MSE) loss function.
- Optimal settings: Batch size of 2 per process, initial learning rate of 0.01, dynamic learning rate adjustment with maximum learning rate of 0.3.



- SGD and Adam optimizers: SGD used with onecycle scheduler and momentum of 0.9.
- Selection of the best model based on validation loss.

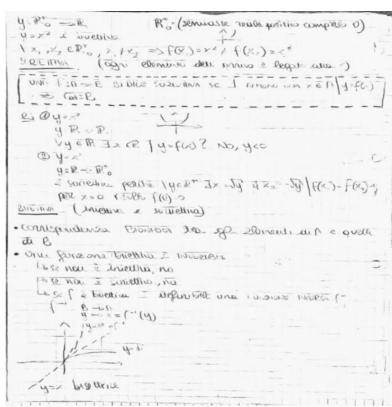


Figure of a real image inferenced

#### **Experimental Methods**

Evaluation metrics used: Mean Squared Error (MSE) and Structural Similarity Index (SSIM).

 MSE: Measures the average of squared differences between the predicted (denoised) and the ground truth images.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left( \hat{I}(i) - I(i) \right)^{2}$$

• SSIM: Compares the similarity between two images, using an uncompressed or distortion-free image as a reference.

SSIM(
$$\hat{I}, I$$
) = 
$$\frac{(2\mu_{\hat{I}}\mu_I + C_1)(2\sigma_{\hat{I}I} + C_2)}{(\mu_{\hat{I}}^2 + \mu_I^2 + C_1)(\sigma_{\hat{I}}^2 + \sigma_I^2 + C_2)}$$

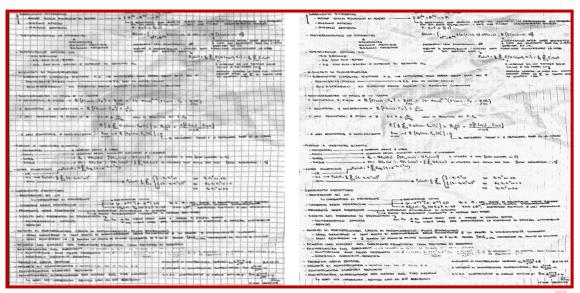


#### **Evaluations and Results**

- The table presents the difference between the denoised images and the ground truth images, contrasted against the synthetically applied grid images and GT.
- Observation: Although evaluation methods like MSE and SSIM provide quantitative results, visual inspection remains the most effective way to assess model efficiency.

- Miles town towns to make the CP AP - P.	- GOALIEUTE STOCKSTOO
- Market attack at a section and section as a section as	1 - normal manuscript to the second to the s
- Minkey Riversity is a second to the second	- CO-STATE CHARGE COLOR AT THE THE THE THE STATE OF THE S
	- Checomo estroguio P. X-4 - D. II
- merupananana sa amanana 22.00 a S <sub>ala a</sub> stalana ababanan - E (kinta-ab. 4)	
Secretaries against the secretaries of a reliable to a reliable	- CONTROLLEMENTO O STANDARDO DE CONTROLLEMENTO D
With Control of the C	BACCURANTEA SURFACED WILL LOSA MINERALIS
SERVICE AND A SERVICE AND A SERVICE AND ASSESSMENT ASSE	DACTORS PROTECTED AND THE PROPERTY OF THE PROPERTY AND TH
The species Suit of B. Electron all - & E. Come	- EDETALEURIA WITTING S.S. SOLO SA BASTIN BATT POSSIBLE (MUNICERALE LA LONG.
	- 20 consists = 2 (( ( ( ( ( ( ( ( ( ( ( ( ( ( ( ( (
- the first state of complete to Market Market Market Complete to Market Market Market Complete to Market M	- 55 800 HIM - 86104
	- SQ COD MICH - SERVEN & HATCHES THE SCANNIC Man (S, NOTED AND C SA TOTAL CO.)
STANDARD DE SERVICIONE DE SERVICIONE DE LA CONTRACTOR COMPANDA MANDE DANS DE CONTRACTOR DE CONTRACTO	
Special Control of the Control of th	- SUSPLEATE OF PORTUGE STATES FOR THE STATES OF THE STATES
Store Sand Alexan to Something worth Industrial of Management (passagement)	- CAUVICEOUT FORTE - E.C. CEA UN UNION WHILE
- DOLD THE REAL PROPERTY CONTRACTOR OF THE PROPERTY OF THE PRO	APPRAISE S REGIST S REGIST OFFICE OFFICE OF A RESIDENCE AND A
	- EDD REGEREOUS CO COMENTO TRANS
SCHOOLSEAN OF WHICH IS 10 HOUSE	- STATUTE CANADA AN HOUSE OF AN ANCIENT
e nomentara. A recon r & [ricon - 5-7 a \$400 - 50 recoj***(\$500 - 70 - \$500)	- + COUNTRIES, & 5-050 - E [FICAL-F.] + BU + (4- MC) (FICAL-F BULL)
of describes, a second-come of Efficient Call to the	Can (Times - To East)
	- F COUNTRIA, & SECRETARION A & FROM - Fo] & D
to your depression, it from a to 1 to the fifty and to their ter to the	
	- F JOH CONVENER, 2 PRICE IS B ! O. R. C. M. LOW SI CHARGOIN SU F. F.
\$ [ 4 \$ + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	E[ - # 11
the set of factors of barts and the second of the second o	E[ # 20 11 9 (com. for 112 ) 6 14 cm + 2 (cm) - Fine)
- a sale adventure, a techniques - time and & [agines, \$_100] and	- 6 MAN CONTENSION, A DECRESCIONES : Sim int & (65 (MM, \$_1)8_1) = 0
	N. TO TOWN O THE P. TOWN A O CONTINUES, THE EL IN THESE
Server & Vancauria, money	
THE PARTY NAMED AND PARTY NAME	-flectour A VARIAUTER QUIDOTTI - grand Q Valend
	A COMPANY DESIGN, MARKETON S. SCHOOL & STREET, S. STREE
ONE STATEMENT OF S	- COURT - & - OF (CO) - (Official) - OF (con) - COURTS & THE (CON MARCO, M. V)
	- SAGE - SEE OF (w) - OF (wg) - 1 2 OF (w) +- UTWEND WE HAVE BOOK THE GOOD SCHOOLS, 1, 4)
man morning policies a filling (no a man man)	- coses remerious - studios, & Bong (4. e - 4. ex to)
- se - tone & the second and the sec	
- ar - second & ( a district, a second ( f. district, as ar district as	- 4 - 4 - 4 - 4 - 4 - 4 - 4 - 4 - 4 - 4
w go I file district per de de de la constance	- que - California de Californ
NO 19 N. N. Tones	a per felandia de prate ma
Management Tab UM	- CEMPROTES ACCULAND
La Contactinate at excessional parameters personal content	- PRODUCENDANI SA) UM
- (u - u - ) LO - u - u - u - u - u - u - u - u - u -	Ly ONDERTORE BY POCKETONE WHITE SQUARE WHITE SQUARE AND SQUARE BY POCKETSCHE WHITE SQUARE
	- TOOLERS SEEM TOOLERS OF THE PARTY OF THE CALL OF THE PARTY OF THE PA
	- PROPRIETA BUILD DIRECTOR - BUILD STANDARD VINEDATO & HERMO WAS S
- bornessenson invitate	
AND	- HOUSE EXPENDED CHITCHED BY THE THE THEORY AND THE EVENTUAL IN THE BUTCHER A PRINCIPLE OF
2001 전 200 100 100 100 100 100 100 100 100 100	-10150
- they benefitue a last south to decrease their benefits of the benefit of the benefit to be the benefit of the	- ANALISI DI BORVETCHINEN (POLITI DI ACCOMUNENDILE) PUNTI STATECHARIO) E UN PRIMO IN STREMINISTICA VINCONTO
- DISCO MACHINE - 2 AMERICA MI STATE MI STATE MACHINE MI STATE MACHINE CONT. THE PROPERTY OF MACHINE M	- CANO GRANE -+ CHAR BONCO AL ACADIMINATION DE CONTRACTOR A CONTRACTOR DE CONTRACTOR D
SCHOOL BELL HETCHED WILL SANCHINEST PROSPERING (Spin Bernstein by Millerton	- ONSO CONTRESO - 2 QUARDO DE PROPO SE FILMO VILLEBATO SE É SU R. ANORE (MET, CAL CANEGORE M. DILVIO DE FINANCE
ACCOMPANIES AND DESCRIPTION OF THE PARTY OF TAXABLE PARTY AND THE PARTY	- SCHENA BEL HETORO WEL GRADUISE PROMITTATO (FOR REINFORD DI NODITTO)
- Realthfree and days the contact - 3 to the total and the contact to the contact to the contact to	- CONVERGENCE SULL CONTROLL - WHITE IN THE ST TO STATE OF THE STATE OF
- Personal Communications and the Communication of	- CONTINUE CHE CRES PROBLEM -> 2 C. (. >> C. 150 C C. 150
TOURS OF SETTING CONTROL STATE OF THE PROPERTY	
AND AND ADDRESS OF THE PARTY OF	- TRUBERA MELA CATTURA - VENETA LES EN C. TURE A C. TURE
	- PATORIA DE PROPERTO DE PROPE
Constitution to the party and the party of t	- CONTROLLEGAR STREETH STREETH SEE DET HERITA DEZ TIPO MERITAN DAVIDE SUPERINGARE DE OUDRE AMERICA DE CANDO DE
Le sont for communical agricult and set self-somment	Le 1900 TOT TOTAL TOTAL LOLD ON PU DE MODERNEY TOTAL ORDER TORDER TOTAL ORDER TOTAL ORDER TOTAL ORDER TOTAL ORDER TOTAL ORDER
of the Importal	

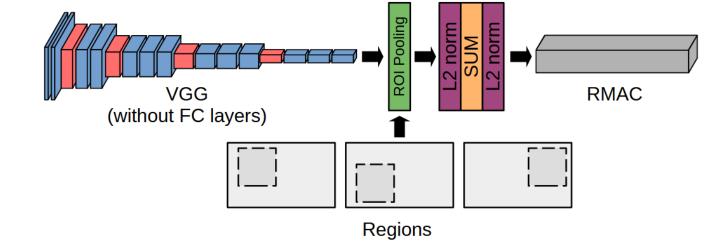
Figure	MSE		SSIM		
	Inferenced	Original	Inferenced	Original	
22	0.0580	0.0063	0.4672	0.6743	
23	0.0691	0.0981	0.4424	0.1724	
24	0.0582	0.0924	0.4745	0.1598	





#### **Image Retrieval Pipeline**

- Input image is preprocessed and converted to tensor
- Tensor is passed through pretrained CNN feature extractor
- 3. Apply RMAC on activations to get regional vectors
- Vectors normalized and summed into global descriptor
- 5. Descriptor compared to database to retrieve top k matches





#### **Computing the RMAC Vector**

MAC vector takes max activation per channel

$$f = [f_1, ..., f_k, ..., f_K]^T$$
, with  $f_k = \max_{x \in X_k} x$ 

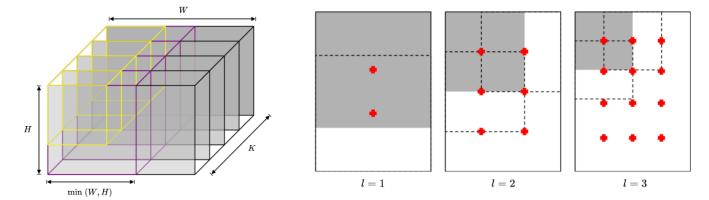
RMAC divides tensor into overlapping regions

$$f_{R_i} = [f_{R_i,1}, ..., f_{R_i,k}, ..., f_{R_i,K}]^T$$
, with  $f_{R_i,k} = \max_{x \in R_{i,k}} x$ 

Regional MAC vector computed per region

$$F = \sum_{i=1}^{N} f_{R_i} = \left[\sum_{i=1}^{N} f_{R_i,1}, ..., \sum_{i=1}^{N} f_{R_i,k}, ..., \sum_{i=1}^{N} f_{R_i,K}\right]^T$$

Vectors normalized, summed to RMAC descriptor



The 3D convolutional activation tensor is divided into overlapping square regions sampled at multiple scales to compute the RMAC descriptor which captures localized features.



#### **Datasets for Image Retrieval**

- Dataset 1 ("db 0") contains 120 unprocessed note images, with 20 images per class, from six different subjects.
- Dataset 2 ("db 1") consists of 120 images, divided into 60 without gridlines and 60 artificially generated images. No connections exist between these two classes.
- Dataset 3 ("db 2") includes 200 images, comprising 20 without gridlines and 180 generated grid images using nine different styles. Associations exist between non-grid and generated grid images across ten classes, with 20 images per class.



#### **Retrieval Performance Evaluation**

 Average Precision (AP) measures relevance of top k retrievals to query

$$AP@k = \frac{1}{k} \sum_{i=1}^{k} \delta(y_i, y_q) \text{ where } \delta(y_i, y_q) = \begin{cases} 1 & \text{if } y_i = y_q \\ 0 & \text{otherwise} \end{cases}$$

mAP calculates the mean AP across all queries

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i@k$$

- Results:
  - Pretrained CNNs achieved highest mAP
  - Demonstrates power of CNN features
  - UNet models performed worse, struggled on complex images
  - Trained UNet had lowest mAP due to overfitting

Table 4: mAP on db\_0

Model	k = 3	k = 5	k = 10	k = 20
VGG16	0.997	0.997	0.993	0.951
VGG19	0.997	0.998	0.994	0.930
DenseNet	1.000	1.000	0.998	0.923
Trained UNet	0.983	0.975	0.864	0.680
Kaiming UNet 0	1.000	0.995	0.989	0.923
Kaiming UNet 1	1.000	1.000	0.998	0.945
Kaiming UNet 2	1.000	0.998	0.996	0.935

Table 5: mAP on db\_1

Model	k = 3	k = 5	k = 10	k = 20
VGG16	0.972	0.955	0.927	0.894
VGG19	0.950	0.937	0.905	0.876
DenseNet	0.972	0.958	0.932	0.894
Trained UNet	0.908	0.858	0.806	0.758
Kaiming UNet 0	0.939	0.903	0.854	0.768
Kaiming UNet 1	0.939	0.912	0.864	0.785
Kaiming UNet 2	0.906	0.868	0.837	0.740

Table 6: mAP on db\_2

Model	k = 3	k = 5	k = 10	k = 20
VGG16	0.675	0.636	0.604	0.548
VGG19	0.717	0.711	0.674	0.594
DenseNet	0.730	0.724	0.677	0.613
Trained UNet	0.467	0.360	0.317	0.302
Kaiming UNet 0	0.467	0.360	0.354	0.348
Kaiming UNet 1	0.467	0.370	0.371	0.375
Kaiming UNet 2	0.467	0.360	0.347	0.336



#### **Some Tests**

- *VGG16* on *db\_0*, *k* = 10
- *VGG16* on *db\_2*, *k* = 20







Top Matches





## Conclusion

- UNet architecture excels in handwritten note image denoising, effectively removing gridlines and noise.
- The pipeline for generating synthetic training data and preprocessing real images shows promise for improving OCR and math symbol detection.
- UNet outperforms the Fourier transform method in gridline denoising.
- Data augmentation and the use of SGD optimization improve model training and performance.
- A retrieval system based on RMAC vectors can identify images requiring denoising.



# Future Development

- The training dataset could be further augmented with more variations to better match real images.
- Instead of training UNet from scratch, fine tune a pretrained model could boost performance.
- Alternative learning rate schedulers like cyclical or one policies could be explored.
- More advanced metrics beyond MSE/SSIM would better evaluate similarity of denoised images to ground truth.
- Using Vision Transformer (ViT) or DINO for image embeddings may improve the retrieval system.
- Retrieval system could be extended to find similarity between crops rather than full images.



# Thanks for the attention cit. Transformer