Document Quality Estimation using Spatial Frequency Response

Pranjal Kumar Rai, Sajal Maheshwari, Vineet Gandhi

IIIT Hyderabad

Presented by : Rajiv Soundararajan

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

CVIT

IEEE

ICASSP CALGARY 2018

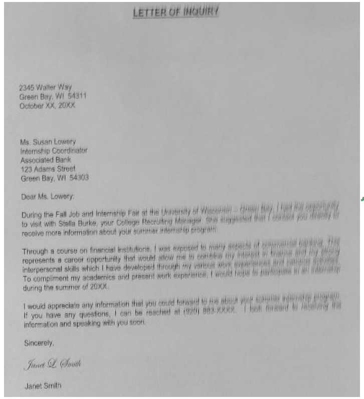
2018 IEEE International Conference on Acoustics, Speech and Signal Processing
15–20 April 2018 • Calgary, Alberta, Canada

Contents

Introduction to Document Image Quality Assessment (DIQA)

- DIQA aims to quantify degradations like focus blur and motion artefacts in a document image.
- Typically, DIQA algorithms can be classified into:
    - **Global DIQA** : Single score for entire image. The score for local regions (if needed) is computed by taking each region as an image.
    - **Local DIQA** : Computes scores for all local regions. All local scores are combined to get a global score.

Introduction to Document Image Quality Assessment (DIQA)



LETTER OF INQUIRY

2345 Walter Way
Green Bay, WI 54311
October XX, 20XX

Ms. Susan Lowery
Internship Coordinator
Associated Bank
123 Adams Street
Green Bay, WI 54303

Dear Ms. Lowery:

During the Fall Job and Internship Fair at the University of Wisconsin – Green Bay, I had the opportunity to visit with Stella Burke, your College Recruiting Manager. She suggested that I contact you directly to receive more information about your summer internship program.

Through a course on financial institutions, I was exposed to many aspects of commercial banking. This represents a career opportunity that would allow me to combine my interest in finance and my strong interpersonal skills which I have developed through my various work experiences and campus activities. To compliment my academics and present work experience, I would hope to participate in an internship during the summer of 20XX.

I would appreciate any information that you could forward to me about your summer internship program. If you have any questions, I can be reached at (920) 883-XXXX. I look forward to receiving the information and speaking with you soon.

Sincerely,

Janet L. Smith

Janet Smith

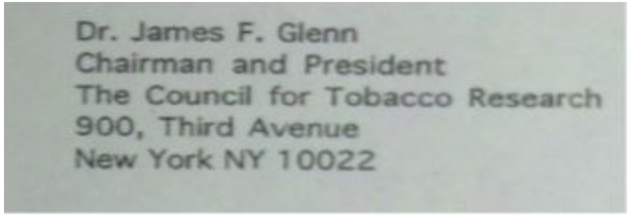
Global DIQA algorithm

Local DIQA algorithm

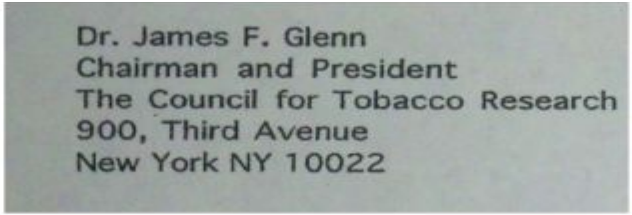
Shortcomings of current DIQA algorithms

A. Dependency on OCR

a. Current algorithms rely on OCR percentage accuracy as ground-truth (Ye et al. [1], Kang et al. [2]).
b. However, OCR accuracies are dependent on font (type, size), language, document layout (presence of tables, graphs etc.), presence of images. Moreover, for identical input, different OCR algorithms can give different results.

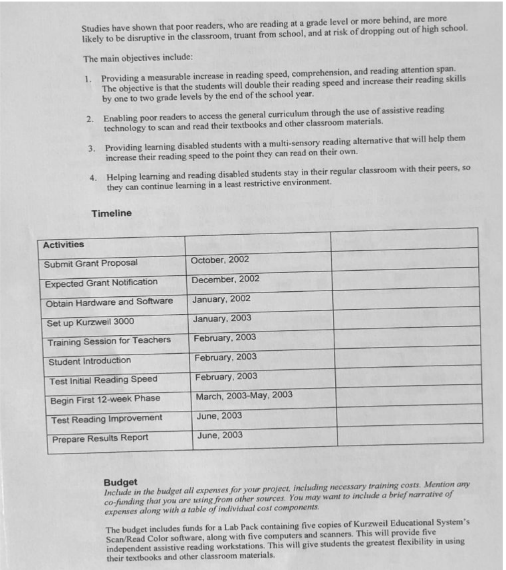

Dr. James F. Glenn
Chairman and President
The Council for Tobacco Research
900, Third Avenue
New York NY 10022

OCR Accuracy: 97.15%


Dr. James F. Glenn
Chairman and President
The Council for Tobacco Research
900, Third Avenue
New York NY 10022

OCR Accuracy: 98.15%

Dependency on OCR examples contd.



Studies have shown that poor readers, who are reading at a grade level or more behind, are more likely to be disruptive in the classroom, truant from school, and at risk of dropping out of high school.

The main objectives include:

1. Providing a measurable increase in reading speed, comprehension, and reading attention span. The objective is that the students will double their reading speed and increase their reading skills by one to two grade levels by the end of the school year.
2. Enabling poor readers to access the general curriculum through the use of assistive reading technology to scan and read their textbooks and other classroom materials.
3. Providing learning disabled students with a multi-sensory reading alternative that will help them increase their reading speed to the point they can read on their own.
4. Helping learning and reading disabled students stay in their regular classroom with their peers, so they can continue learning in a least restrictive environment.

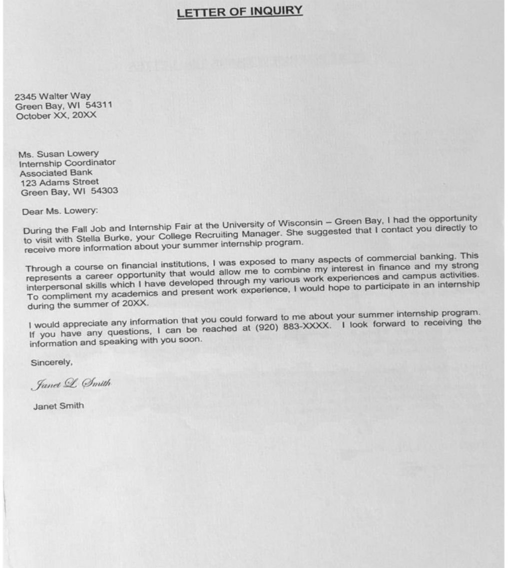
Timeline

Activities		
Submit Grant Proposal	October, 2002	
Expected Grant Notification	December, 2002	
Obtain Hardware and Software	January, 2002	
Set up Kurzweil 3000	January, 2003	
Training Session for Teachers	February, 2003	
Student Introduction	February, 2003	
Test Initial Reading Speed	February, 2003	
Begin First 12-week Phase	March, 2003-May, 2003	
Test Reading Improvement	June, 2003	
Prepare Results Report	June, 2003	

Budget
Include in the budget all expenses for your project, including necessary training costs. Mention any co-funding that you are using from other sources. You may want to include a brief narrative of expenses along with a table of individual cost components.

The budget includes funds for a Lab Pack containing five copies of Kurzweil Educational System's Scan/Read Color software, along with five computers and scanners. This will provide five independent assistive reading workstations. This will give students the greatest flexibility in using their textbooks and other classroom materials.

OCR Accuracy: 78.09%



LETTER OF INQUIRY

2345 Walter Way
Green Bay, WI 54311
October XX, 20XX

Ms. Susan Lowery
Internship Coordinator
Associated Bank
123 Adams Street
Green Bay, WI 54303

Dear Ms. Lowery:

During the Fall Job and Internship Fair at the University of Wisconsin – Green Bay, I had the opportunity to visit with Stella Burke, your College Recruiting Manager. She suggested that I contact you directly to receive more information about your summer internship program.

Through a course on financial institutions, I was exposed to many aspects of commercial banking. This represents a career opportunity that would allow me to combine my interest in finance and my strong interpersonal skills which I have developed through my various work experiences and campus activities. To compliment my academics and present work experience, I would hope to participate in an internship during the summer of 20XX.

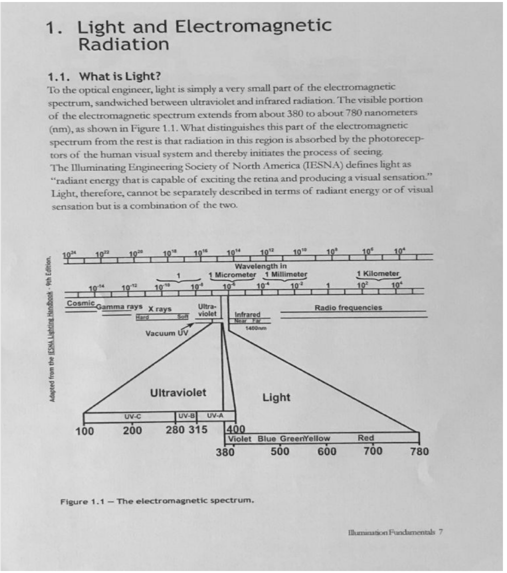
I would appreciate any information that you could forward to me about your summer internship program. If you have any questions, I can be reached at (920) 883-XXXX. I look forward to receiving the information and speaking with you soon.

Sincerely,

Janet L. Smith

Janet Smith

OCR Accuracy: 47.94%



1. Light and Electromagnetic Radiation

1.1. What is Light?

To the optical engineer, light is simply a very small part of the electromagnetic spectrum, sandwiched between ultraviolet and infrared radiation. The visible portion of the electromagnetic spectrum extends from about 380 to about 780 nanometers (nm), as shown in Figure 1.1. What distinguishes this part of the electromagnetic spectrum from the rest is that radiation in this region is absorbed by the photoreceptors of the human visual system and thereby initiates the process of seeing.
The Illuminating Engineering Society of North America (IESNA) defines light as "radiant energy that is capable of exciting the retina and producing a visual sensation." Light, therefore, cannot be separately described in terms of radiant energy or of visual sensation but is a combination of the two.

Figure 1.1 – The electromagnetic spectrum.

Illumination Fundamentals 7

OCR Accuracy: 72.50%

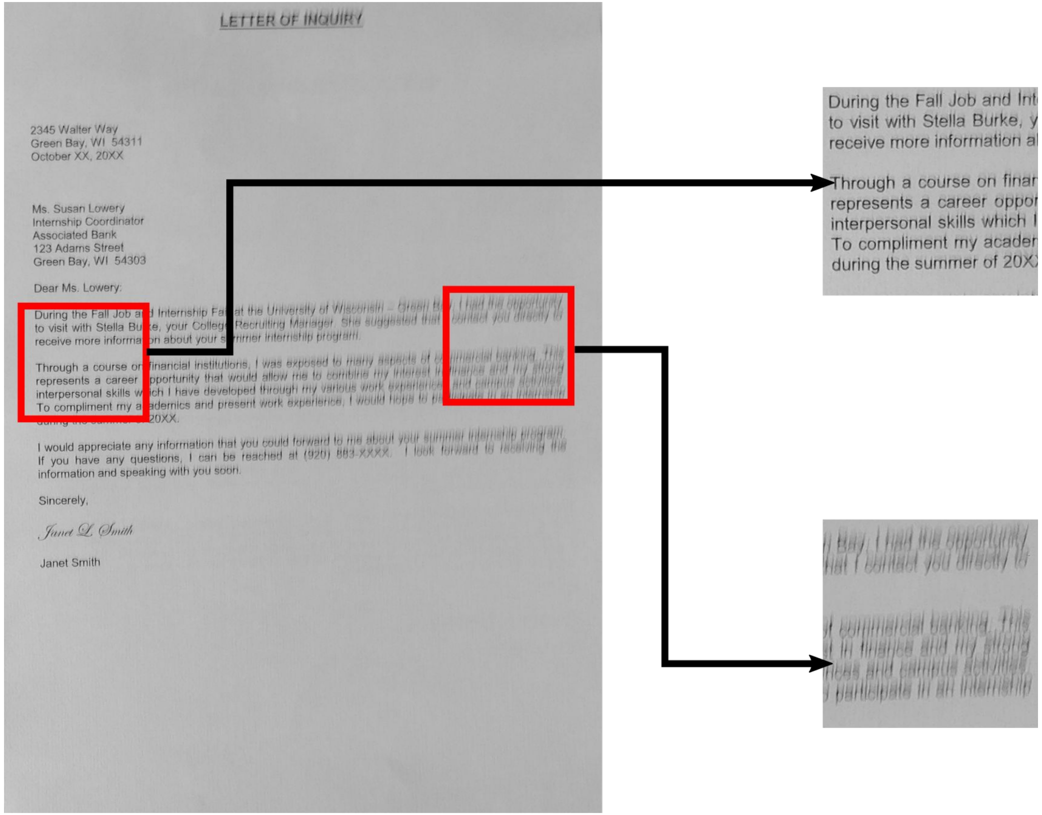
The above examples clearly demonstrate that OCR accuracies are not invariant towards presence of tables, layout changes and diagrams.

Shortcomings of current DIQA algorithms

B. Erroneous intra-document representation

a. Most current state-of-the-art algorithms use a single ground truth value for input images (Ye et al. [1], Kang et al. [2]).
b. The single ground truth approach is correct only if:
    i. The entire image is uniform with no intra-document variations, which is rarely the case.
    ii. The decision is based on a certain region of the document, e.g. the best/worst part of the input.
c. Using a single ground truth for entire image leads to erroneous training and incorrect evaluation.

Erroneous intra-document representation example

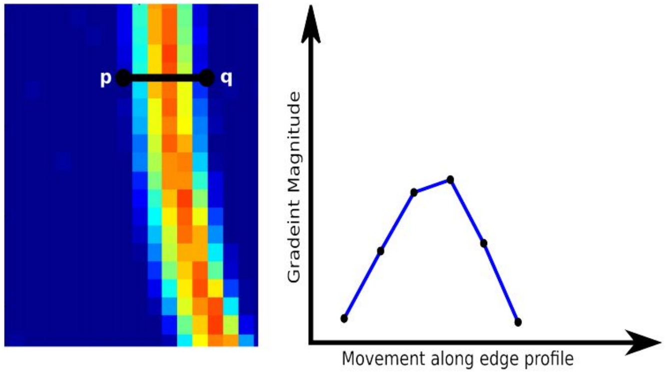


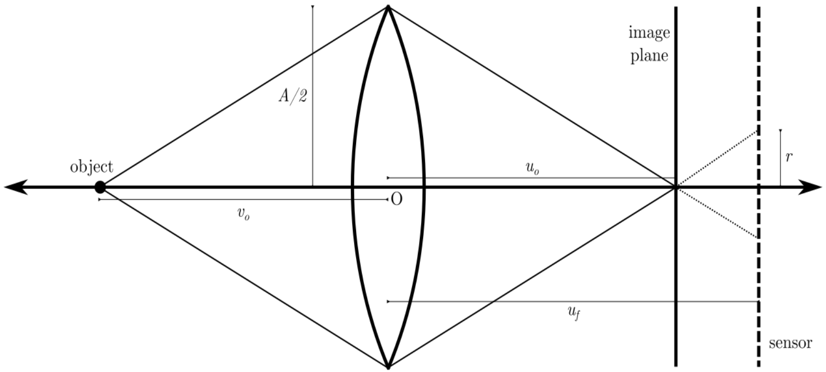
This representation shows two regions of the same input image.

The large intra-document difference introduced due to motion effects can be seen in the diagram.

A single value of ground truth such as OCR accuracy can not capture these differences.

Other recent attempts (limited to focus blur)





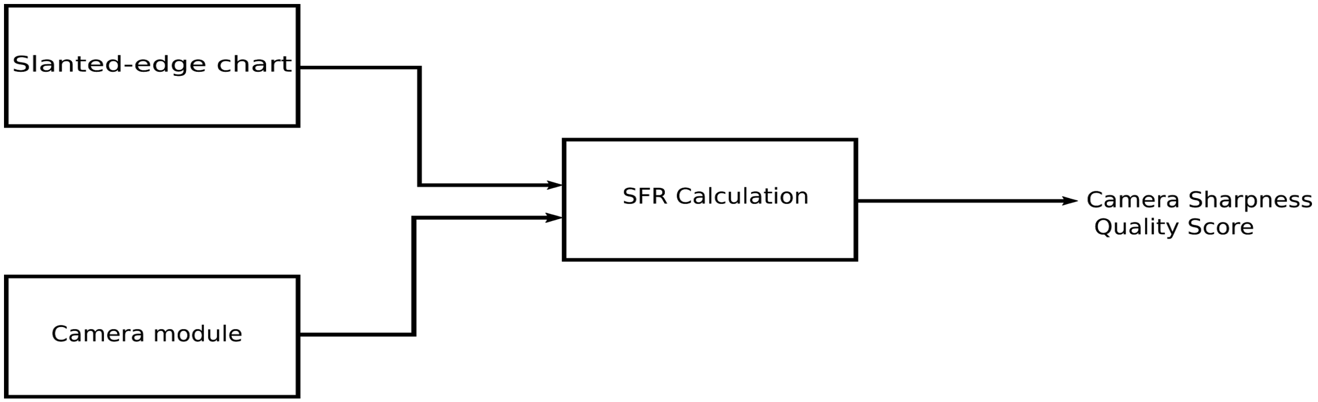
- Maheshwari et al. ICVGIP’16
- Analyze transitions between text and non text region

- Rai et al. ICDAR’17
- Computationally vary the camera focus distance to calculate radius of blur ground truth

These approaches are limited to out-of-focus blur and cannot handle motion artefacts.

Proposed solution : Spatial Frequency Response (SFR)

- We propose an alternate way to calculate ground truth which can be used for both local and global DIQA.
- The proposed method uses the concept of Spatial Frequency Response, traditionally used to determine camera quality in terms of sharpness.



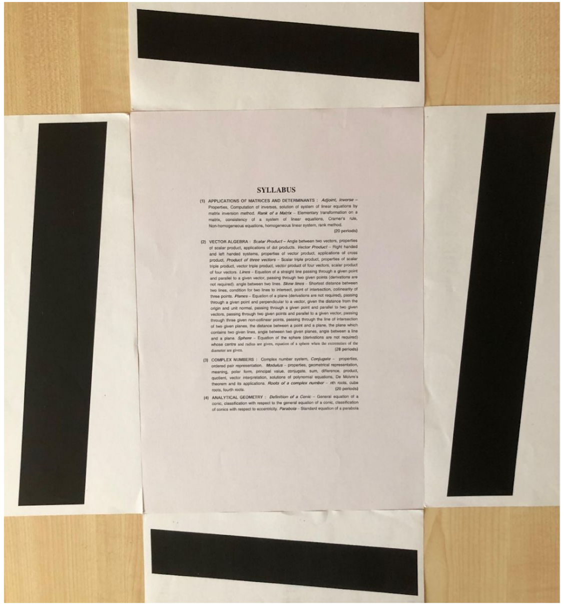
SFR calculation schematic to predict sharpness of smartphone cameras.

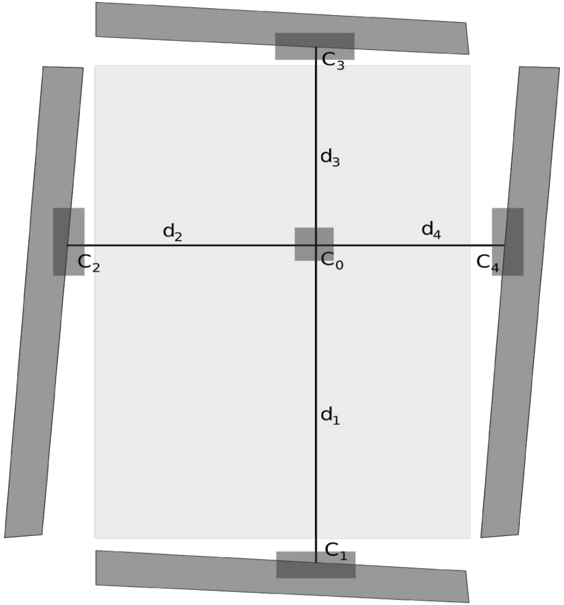
Proposed solution : Spatial Frequency Response (SFR)

- Our main insight is that if the camera module is not changed and the quality score changes, then that change is due to the difference in input image quality.
- We apply this insight on our training images, with a setup to generate local quality scores using slanted-edges technique, typically used for SFR.
- Using our ground-truth and training images, we train a Convolutional Neural Network (CNN). This can be used for quality estimation of document images (locally and globally) in uncontrolled settings.

Proposed solution : Spatial Frequency Response (SFR)

- The setup for the calculation of local SFR scores and a schematic representation for score calculation is shown below.





Proposed solution : Spatial Frequency Response (SFR)

- We calculate the SFR score for each ‘*slanted-edge patch*’ corresponding to the patch in the input document image.
- We calculate a final score as the weighted mean of two patches with minimum scores, weighted according to the distance from the patch.
- We have generated an extensive dataset using the above stated guidelines which we refer to as the SFR dataset.

Calculation of SFR values

- Traditionally, SFR is used to calculate the '*Resolution Power*' of a camera.
- For SFR calculation, we require an input with all frequencies, i.e. a step function.
- However, for SFR to work, we need frequency from the sensor beyond its pitch. This is why we use the *'slanted-edge'* technique.
- Since the value of intensity remains constant along the edge and gradually decreases along the direction of the gradient, a slanted edge helps to capture intensity variations less than the sensor dimensions.

Calculation of SFR values

- The Fourier Transform of array of intensity values across gradient axis therefore, gives Spatial Frequencies much higher than Nyquist limit of sensor dimensions.
- The ratio of the above Fourier Transform with DC component of the signal gives the resolution ratio at all frequencies.
- If the resolution ratio falls below a certain threshold (typically equal to 0.5) for a frequency, that frequency is the SFR score for the particular camera assembly.

Quality Estimation Pipeline

- Patch Extraction
- Ground truth calculation
- Training & Testing

Quality Estimation Pipeline : Patch Extraction

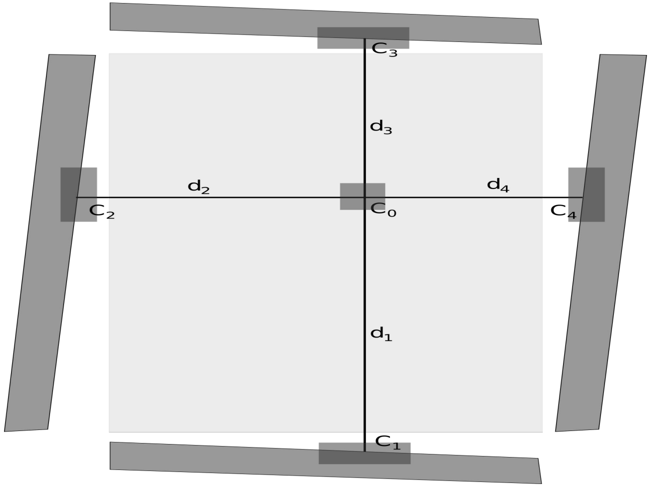
- Non document regions in the input image are removed using segmentation of bounding box with largest connected component, assuming that the document occupies the largest part of the image (Ye et al. [1]).
- Traditionally, DIQA approaches used either random or binarization based patch selection.
- Random patch selection leads to selection of non-informative background patches.
- Binarization algorithms get affected with presence of blur in image.
- We use the intelligent informative patch extraction as proposed in Rai et al. [4].
    - Patches are centred on the transition between text and non-text region.
    - These transition edge profiles are prominent indicators of quality in image.

Quality Estimation Pipeline : Patch Extraction

Comparison of patches selected using binarization and our approach for a low quality image

Quality Estimation Pipeline : Ground truth calculation

- We use the '*slanted-edge*' technique described before to calculate SFR value for each selected patch.
- Patches from the same document image can have different ground truth value in case of intra-document quality variation.



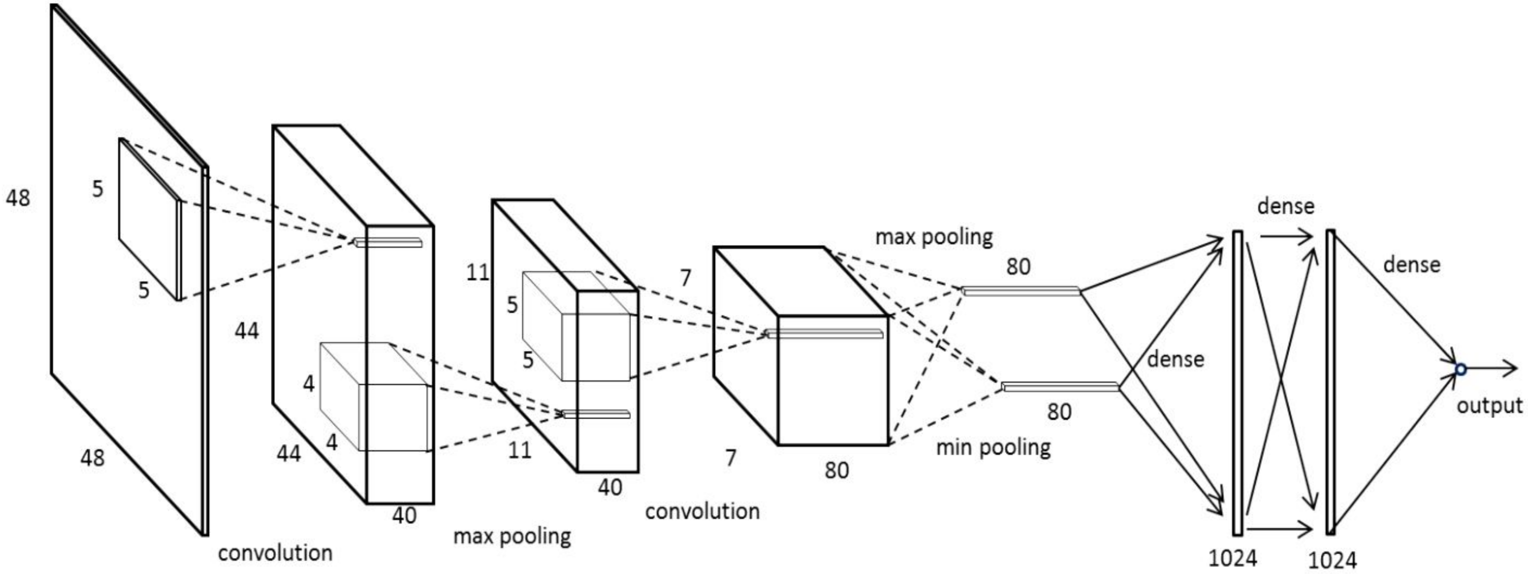
$$g = \frac{\frac{s_i}{d_i} + \frac{s_j}{d_j}}{\frac{1}{d_i} + \frac{1}{d_j}}$$

s_i is SFR value at c_i

Quality Estimation Pipeline : Training & Testing

- The selected patches along with their ground truth values are used to train a Convolutional Neural Network proposed in Kang et al. [2].
- The trained model is used to predict the quality score for test patches, selected from the test image in the same fashion.
- We generate a quality-map and a global quality score for the entire image as the mean of all local scores using the pipeline.

Quality Estimation Pipeline : Training & Testing



CNN Architecture

Results and Comparisons

- We present the results of various learning and non-learning based DIQA algorithms on the proposed dataset.
- We compare the algorithms using both OCR and the proposed ground truth.
- The metrics used for comparison are:
  - Linear Cross Correlation: Measures the degree of similarity between two curves.
  - Spearman Rank Order Cross Correlation: Measures the similarity of monotonicity between two curves.
- The metrics have been applied for both global and local DIQA.

Results and Comparisons

	LCC	SROCC
ΔDOM	0.64	0.65
FOCUS MEASURE	0.69	0.80
CORNIA	0.89	0.87
EPM	0.79	0.82
DCNN	0.85	0.82

Comparison of different approaches on SFR Dataset with OCR accuracies as ground truth

Results and Comparisons

	LCC	SROCC
ΔDOM	0.80	0.74
FOCUS MEASURE	0.70	0.89
CORNIA	0.96	0.87
EPM	0.94	0.89
PROPOSED APPROACH	0.97	0.89

Comparison of different approaches on SFR Dataset with proposed ground truth

Results and Comparisons

- As can be seen from the results above, the proposed ground truth increases the accuracies for all algorithms.
- Also, on comparing the local DIQA scores, training the network with the proposed ground truth leads to a substantial increase in the prediction accuracy over using OCR as ground truth (For DCNN approach, LCC and SROCC for local patches increase from 0.74 to 0.90 and 0.73 to 0.83 respectively).

Conclusions and Future Work

- We have used the concept of SFR to quantify the quality of a document image at patch level.
- Using extensive experiments we have demonstrated that the proposed ground truth leads to a more accurate training of deep neural networks.
- We plan to explore this dataset for the application of deblurring document images.

References

1. Peng Ye, Jayant Kumar, Le Kang, and David Doermann, “Unsupervised feature learning framework for no-reference image quality assessment,” in CVPR, 2012.
2. Le Kang, Peng Ye, Yi Li, and David Doermann, “A deep learning approach to document image quality assessment,” in ICIP, 2014.
3. Sajal Maheshwari, Pranjal Kumar Rai, Gopal Sharma, and Vineet Gandhi, “Document blur detection using edge profile mining,” in Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing. ACM, 2016, p. 23.
4. Pranjal Kumar Rai, Sajal Maheshwari, Ishit Mehta, Parikshit Sakurikar, and Vineet Gandhi, “Beyond ocrs for document blur estimation,” 2017.
5. Marc¸al Rusinol, Joseph Chazalon, and Jean-Marc Ogier, “Combining focus measure operators to predict ocr accuracy in mobile-captured document images,” in Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on, 2014, pp. 181–185.

Questions ?

For any queries and access of dataset, please drop an email to:

pranjal.kumarrai@research.iiit.ac.in

sajal.maheshwari@alumni.iiit.ac.in

vgandhi@iiit.ac.in

Thank You