Small, Medium, and Big Data: Application of Machine Learning Methods to the Solution of Real-World Imaging and Printing Problems

A Personal Journey

Jan Allebach
Electronic Imaging Systems Laboratory (EISL)
Purdue University

1 May 2018

PURDUE UNIVERSITY

What are the Essential Ingredients of Machine Learning? (1/2)

- A well-defined task
  - Choose a decision from a finite set of outcomes, based on observed data.
  - Estimate or predict the value of a continuous variable, based on observed data.
- A well-defined decision or estimation structure
  - Clustering
  - Decision tree
  - Linear regression
  - Support vector machine
  - Neural network, including convolutional neural network (CNN)
  - Or other

What are the Essential Ingredients of Machine Learning? (2/2)

- Features
  - Computed from observed data.
  - Serve as input to the decision or estimation structure.
  - May be handcrafted or determined autonomously as part of the training process.
- Training data
  - Representative of the observed data.
  - Sufficiently diverse or rich to avoid over-fitting.
- A well-defined cost function to penalize errors in classification or estimation.
- A procedure for training the free parameters of the decision or estimation structure to minimize the cost function.

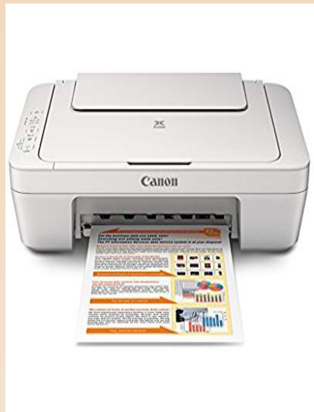
Synopsis

- K Nearest Neighbor classification applied to printer forensics
- Extension of K Means to Scalar Sequential Quantization
- Optimal tree-structured piece-wise linear filter for image scaling
- Training-based methods for digital haftoning
- Black-box model for print prediction based on training and linear regression
- Print macrouniformity prediction (Method 1)
- Print macrouniformity prediction (Method 2)
- Fashion photograph aesthetic quality predictor based on SVM and CNN
- Facial landmark detection using CNN
- Logo identification using CNN
- Text field category classification via natural language processing

Printer Forensics

Whodunnit?

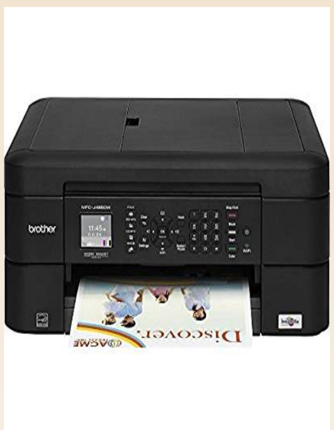
HP Deskjet 1112



Canon MG2522

HP Deskjet 2655

Epson XP-340



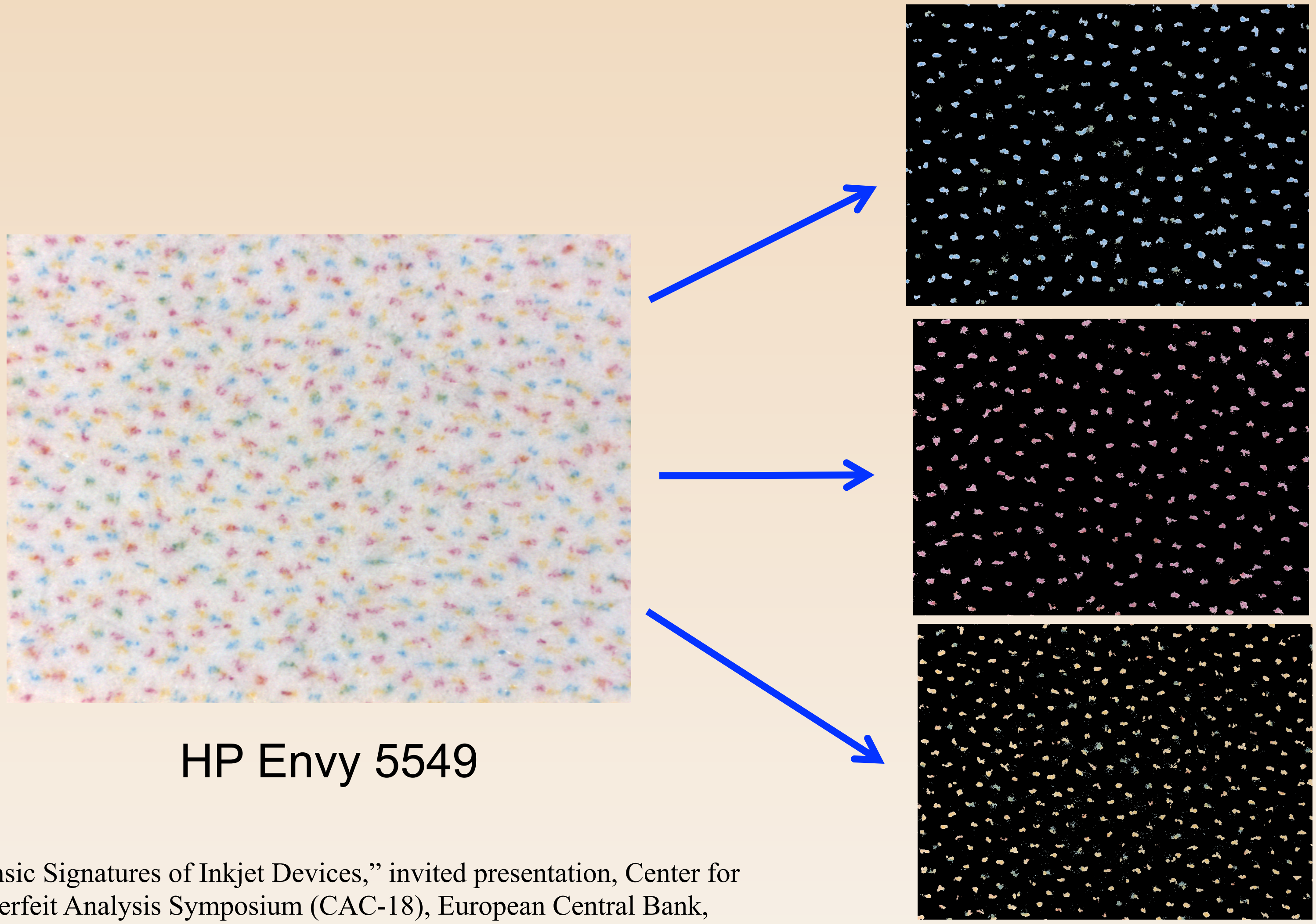
Brother MFC-J485DW

HP Envy 5549

Canon MX922

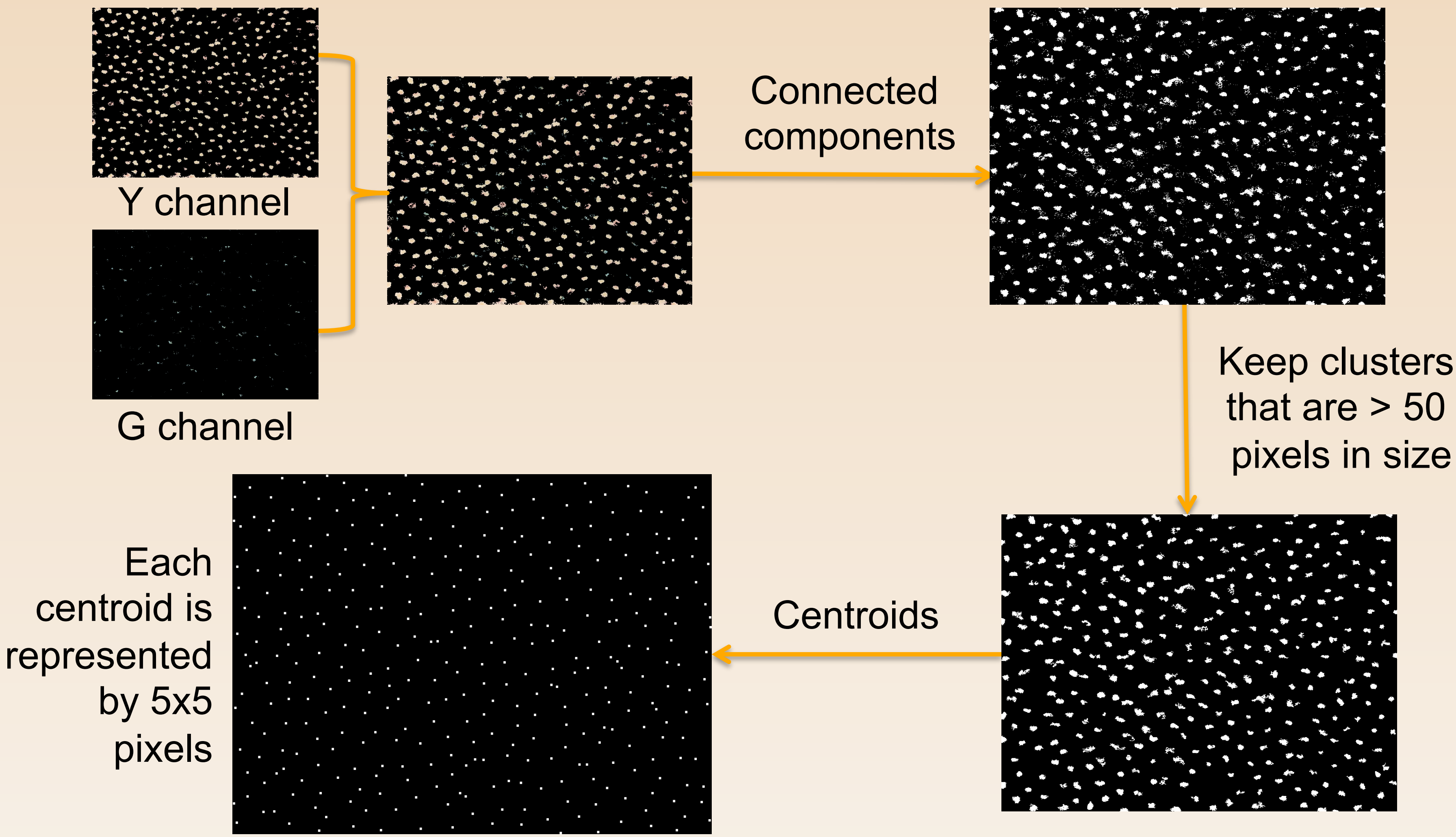
Canon PIXMA MG3620

Supervised Clustering
K Nearest Neighbors (KNN)



"Intrinsic Signatures of Inkjet Devices," invited presentation, Center for Counterfeit Analysis Symposium (CAC-18), European Central Bank, Frankfurt Am Main, Germany, 6-7 March 2018.

Example image analysis for HP Envy 5549 Y and G clusters



Unsupervised Clustering
K-means

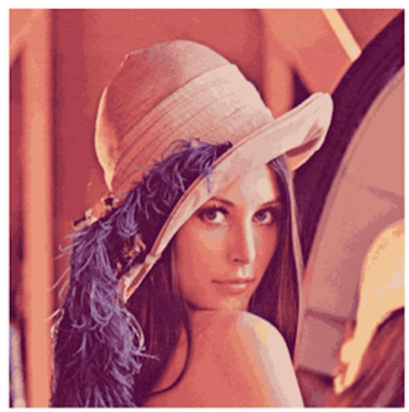
Original

2 Clusters

3 Clusters

6 Clusters

50 Clusters →



A special case of K-means: Structured Vector Quantization*

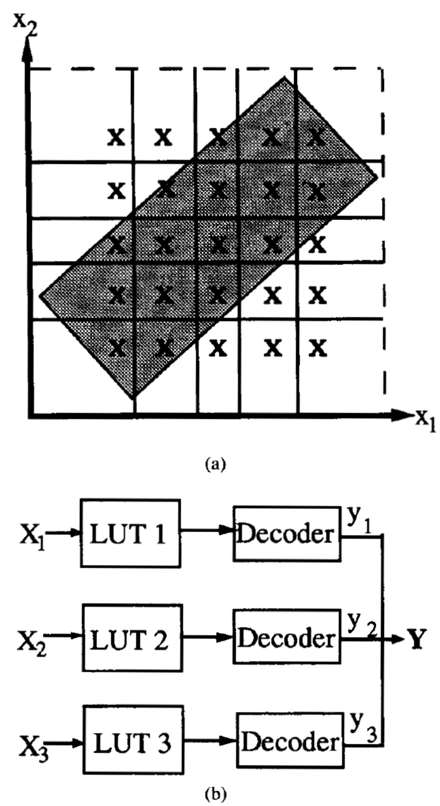


Fig. 1. (a) Two-dimensional example and (b) encoder-decoder operation in independent scalar quantization.

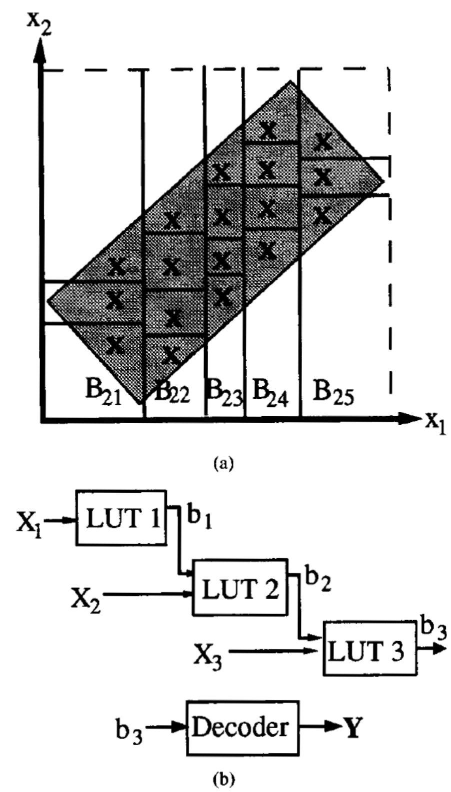


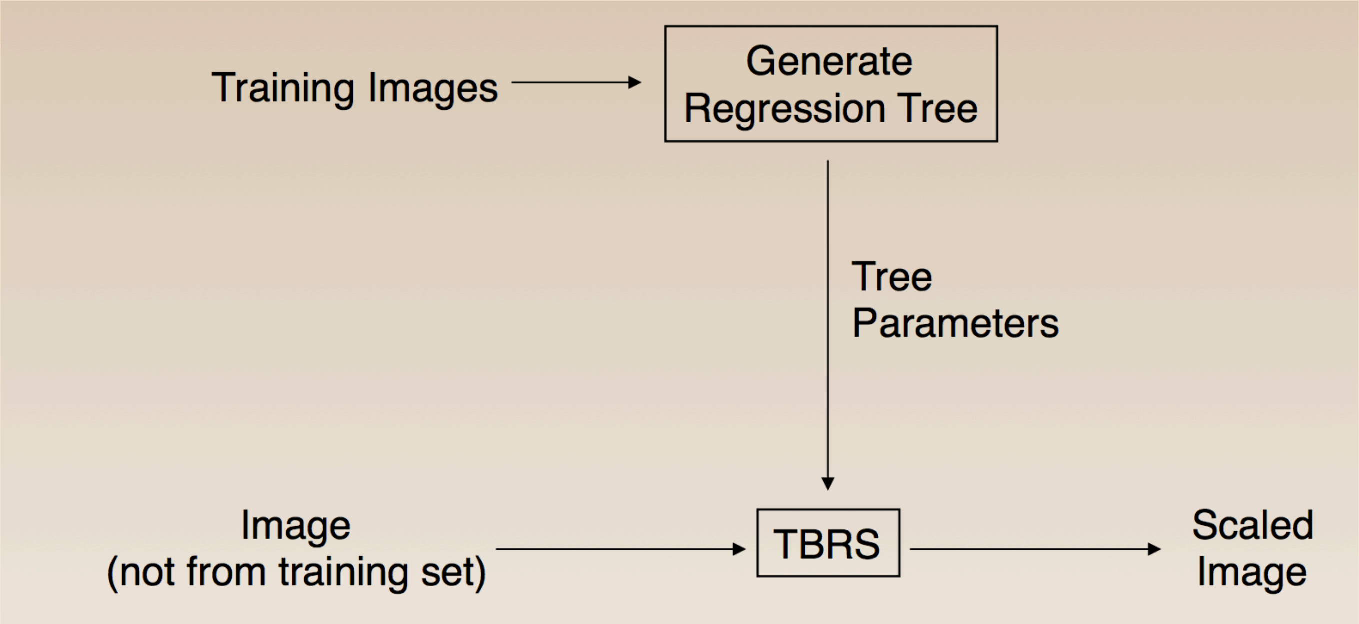
Fig. 2. (a) Two-dimensional example and (b) encoder-decoder operation in sequential scalar quantization.

*Research supported by Eastman Kodak Company.

R. Balasubramanian, C. A. Bouman, and J. P. Allebach, "Sequential Scalar Quantization of Vectors: An Analysis," *IEEE Trans. on Image Processing,* Vol. 4, pp. 1282-1295, September 1995.

J. Z. Chang, J. P. Allebach, and C. A. Bouman, "Sequential Linear Interpolation of Multidimensional Functions," *IEEE Trans. on Image Processing*, Vol. 6, pp. 1231-1245, September 1997.

Tree-Structured Classifiers: Resolution Synthesis – An Optimal Piecewise Linear Interpolator*



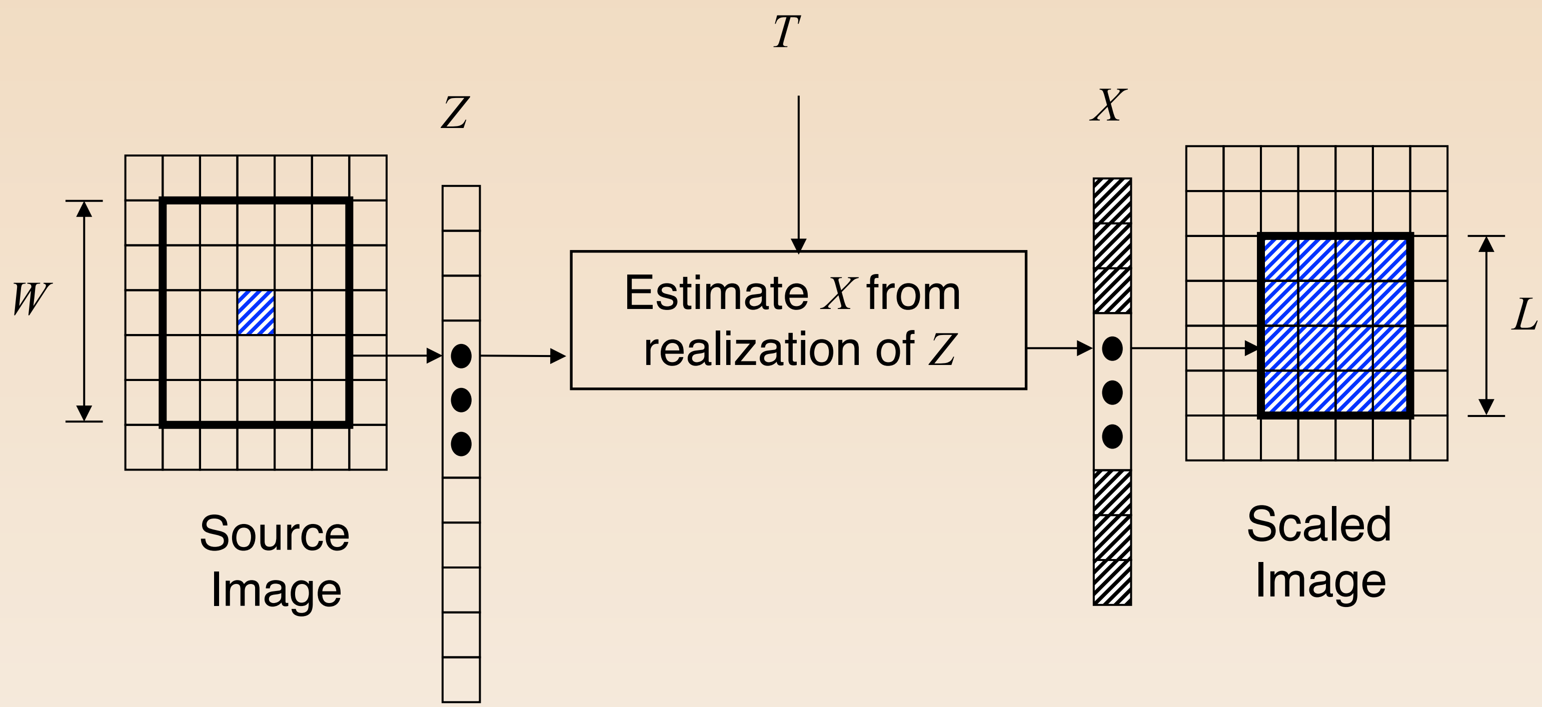
C. B. Atkins, C. A. Bouman, and J. P. Allebach, "Tree-Based Resolution Synthesis," *Proceedings of PICS-99: the 1999 IS&T Image Processing, Image Quality, Image Capture Systems Conference*, Savannah, GA, 25-28 April 1999.

C. B. Atkins, C. A. Bouman, and J. P. Allebach, "Optimal Image Scaling Using Pixel Classification," *Proceedings of the 2001 International Conference on Image Processing*, Thessaloniki, Greece, 7 October – 10 October 2001.

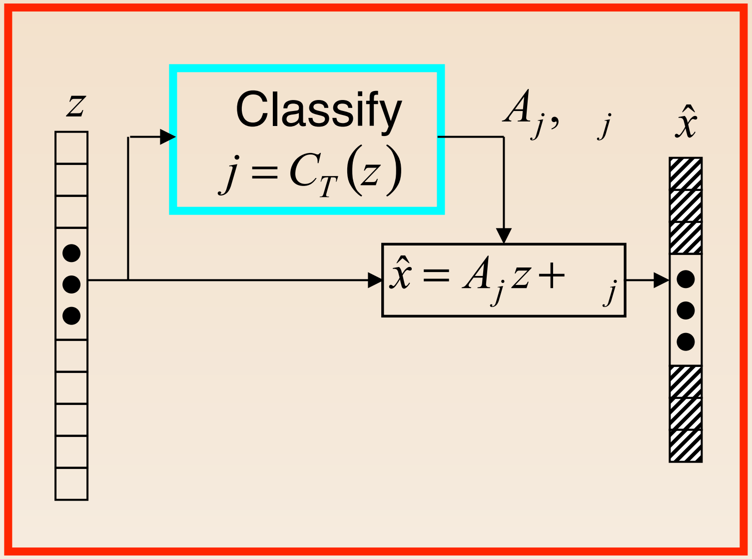
B. Zhang, J. P. Allebach, J. Gondek, and M. Schramm, "Improved Resolution Synthesis Algorithm for Image Interpolation," *Proceedings of NIP22 22nd International Conference on Digital Printing Technologies*, Denver, CO, 17-22 September 2006.

*Research supported by HP, Inc.

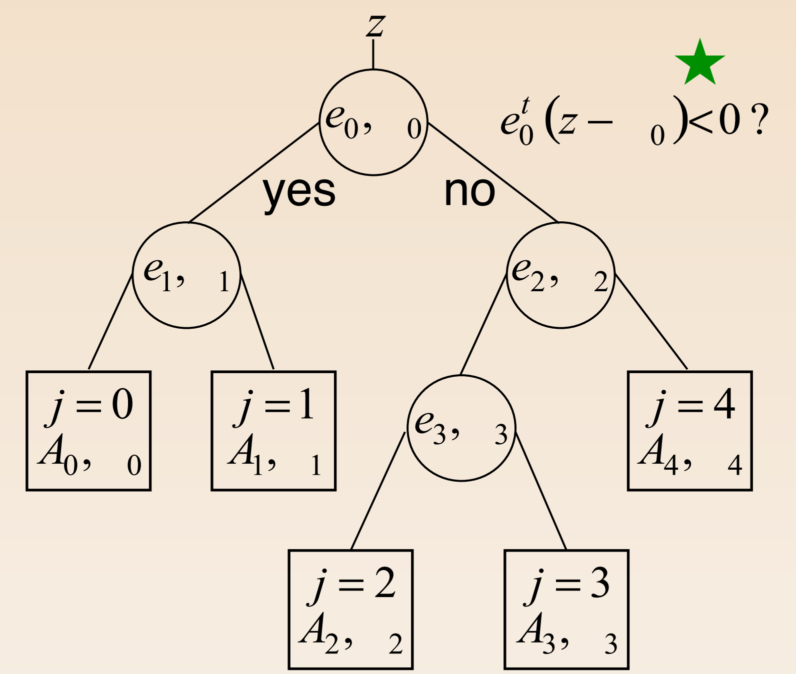
Optimal image scaling



Scaling procedure



$$C_T(\cdot): Z \rightarrow \{0, \ldots, M-1\}$$



4X scaling results

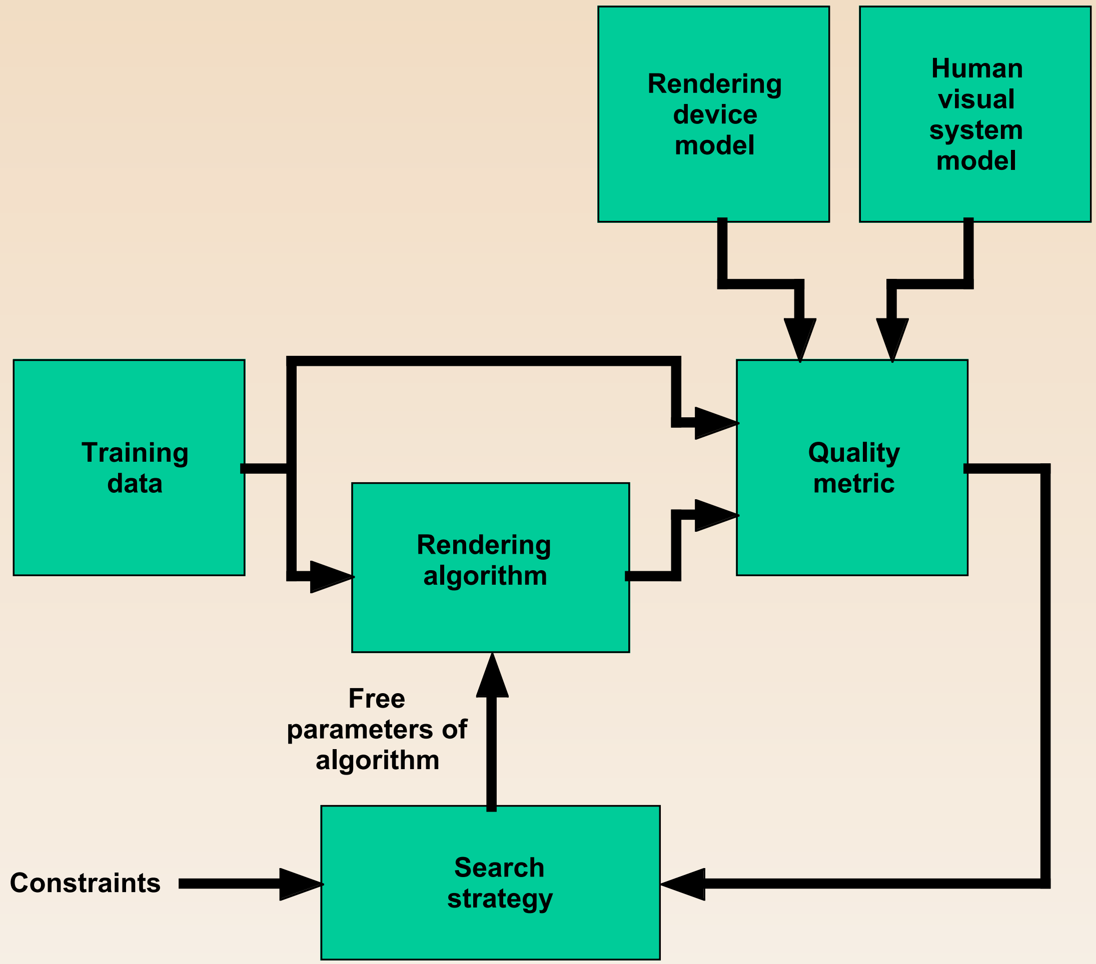
Photoshop Bicubic Interpolation

Tree-Based Resolution Synthesis

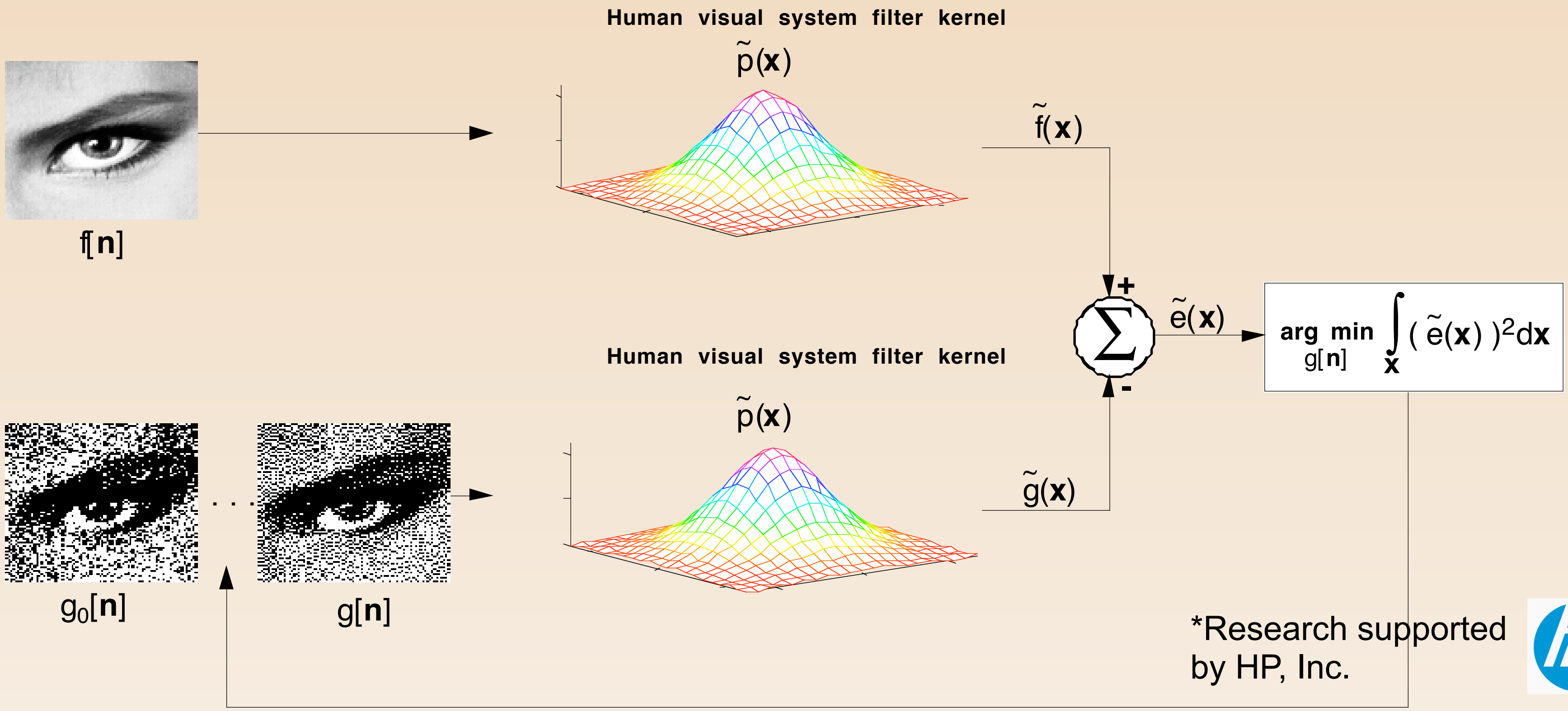
Synopsis

- K Nearest Neighbor classification applied to printer forensics
- Extension of K Means to Scalar Sequential Quantization
- Optimal tree-structured piece-wise linear filter for image scaling
- Training-based methods for digital haftoning
- Black-box model for print prediction based on training and linear regression
- Print macrouniformity prediction (Method 1)
- Print macrouniformity prediction (Method 2)
- Fashion photograph aesthetic quality predictor based on SVM and CNN
- Facial landmark detection using CNN
- Logo identification using CNN
- Text field category classification via natural language processing

Training-based development of optimal rendering algorithms



Model-Based Halftoning: Direct Binary Search (DBS)*

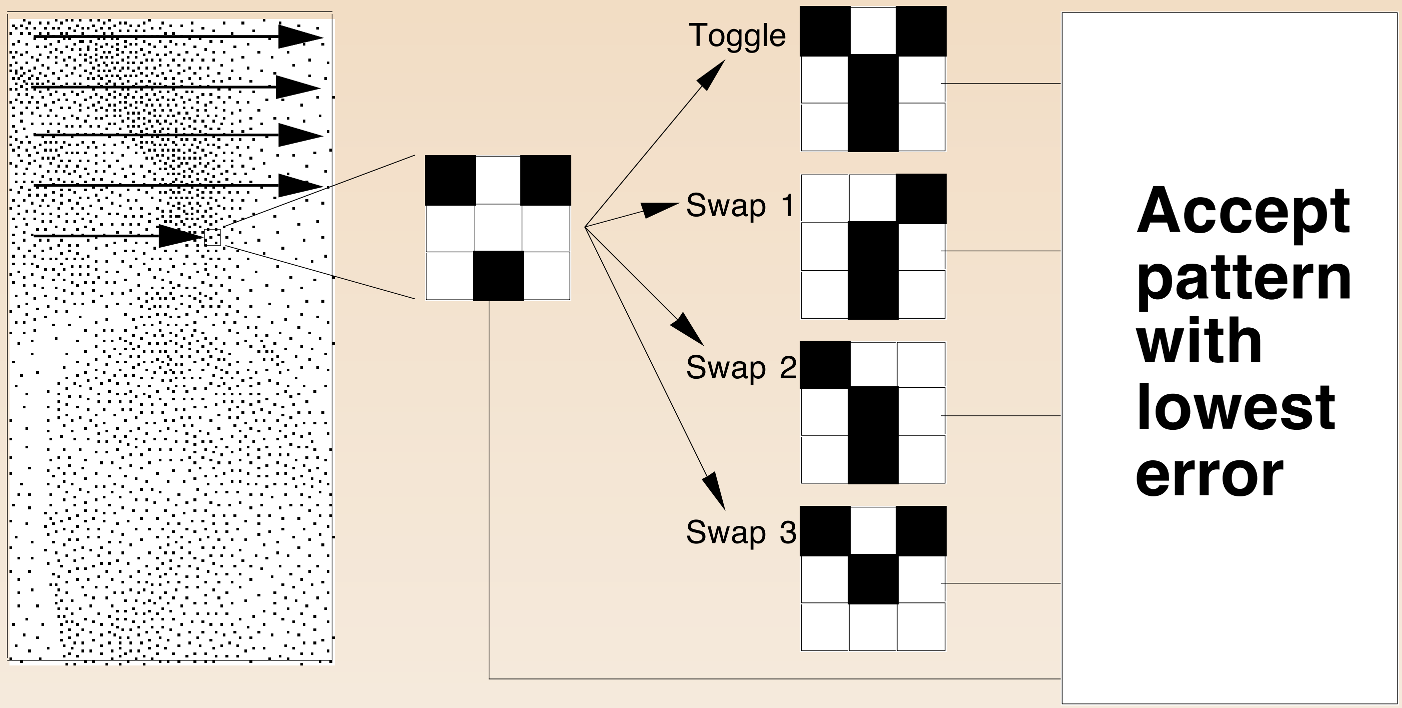


*Research supported by HP, Inc.

Analoui and J. P. Allebach, "Model-based Halftoning by Direct Binary Search," *Proceedings of the 1992 SPIE/IS&T Symposium on Electronic Imaging Science and Technology*, San Jose, CA, February 9-14, 1992, Vol. 1666, pp. 96-108.

D. J. Lieberman, and J. P. Allebach, "A Dual Interpretation for Direct Binary Search and its Implications for Tone Reproduction and Texture Quality," *IEEE Trans. on Image Processing*, Vol. 9, pp. 1950-1963, November 2000.

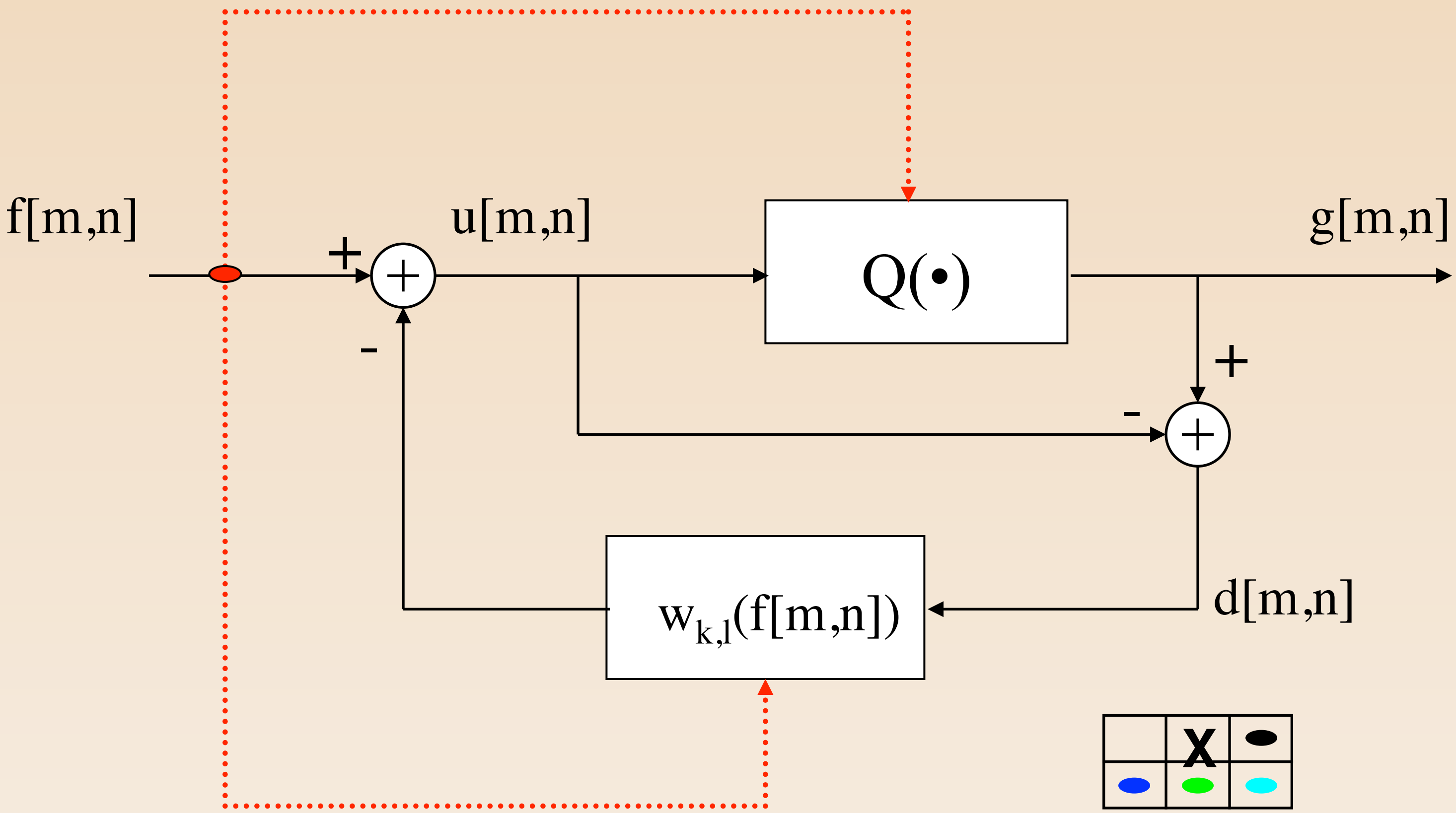
The DBS search heuristic



PURDUE UNIVERSITY

DBS convergence: 0, 1, 2, 4, 6, and 8 iterations

Model-Based Training Supervised Halftoning Tone-Dependent Error Diffusion (TDED)*

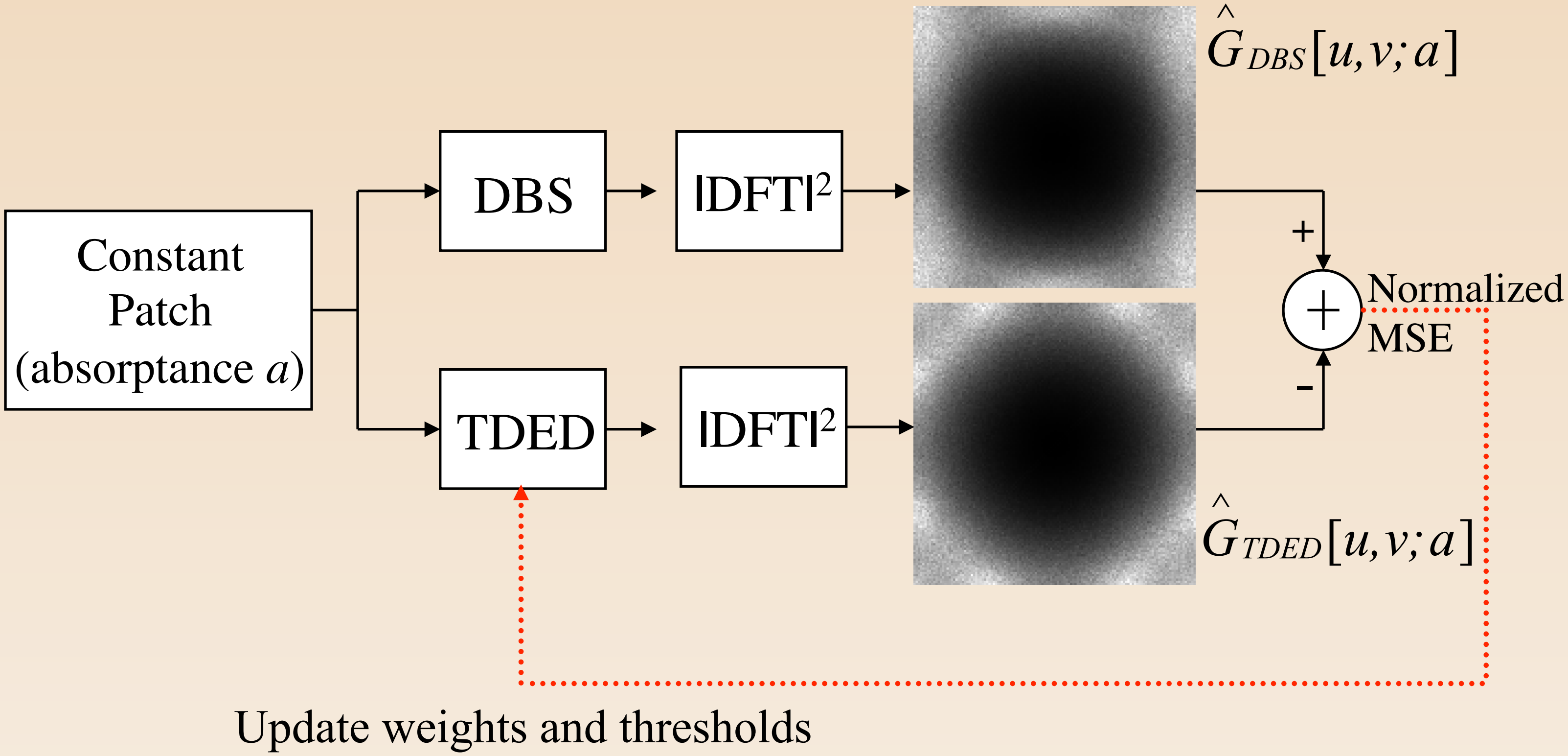


P. Li and J. P. Allebach, "Tone-Dependent Error Diffusion," *IEEE Trans. on Image Processing*, Vol. 13, pp. 201-215, February 2004.

*Research supported by HP, Inc.

PURDUE UNIVERSITY

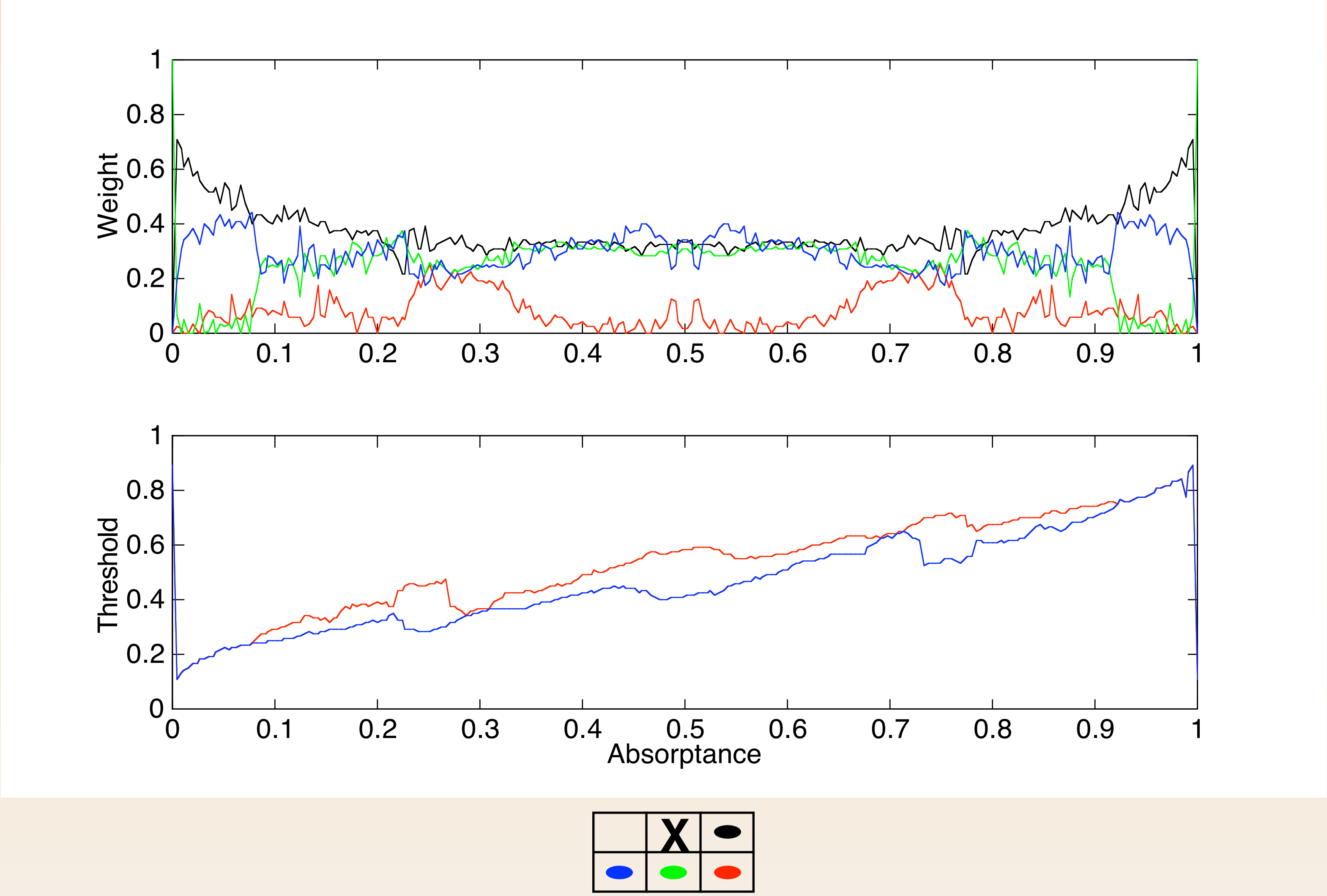
Optimization of TDED parameters



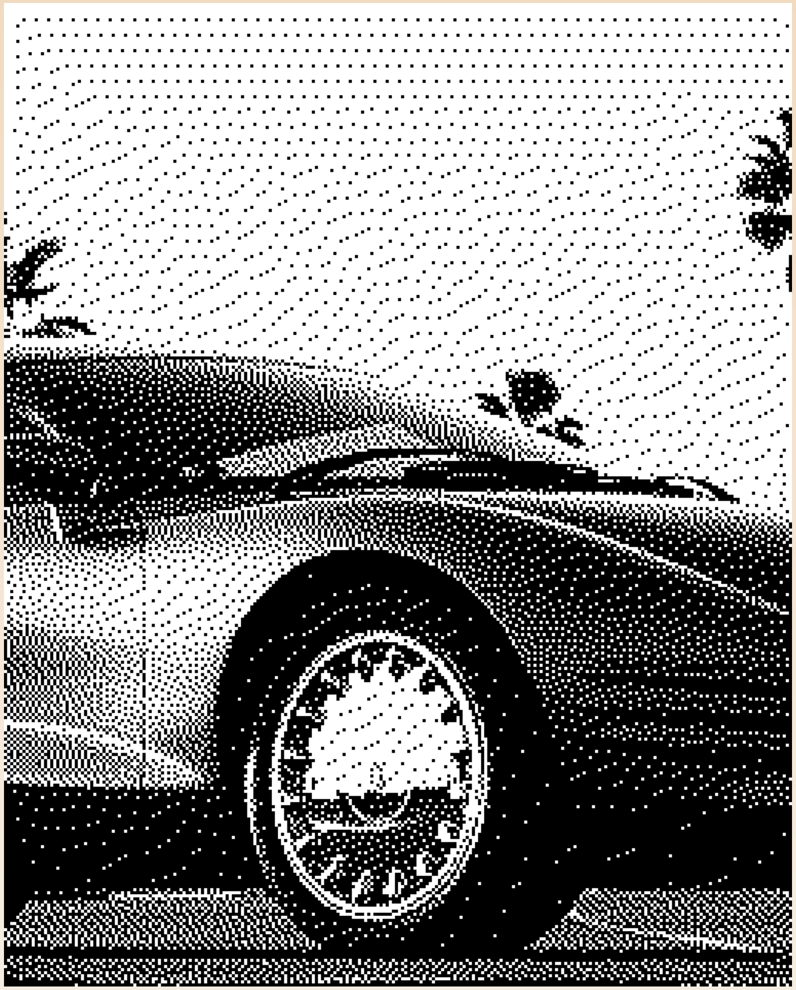
- Cost function

$$\xi(a) = \sum_u \sum_v \frac{(\hat{G}_{DBS}[u,v;a] - \hat{G}_{TDED}[u,v;a])^2}{\hat{G}_{DBS}[u,v;a]^2}.$$

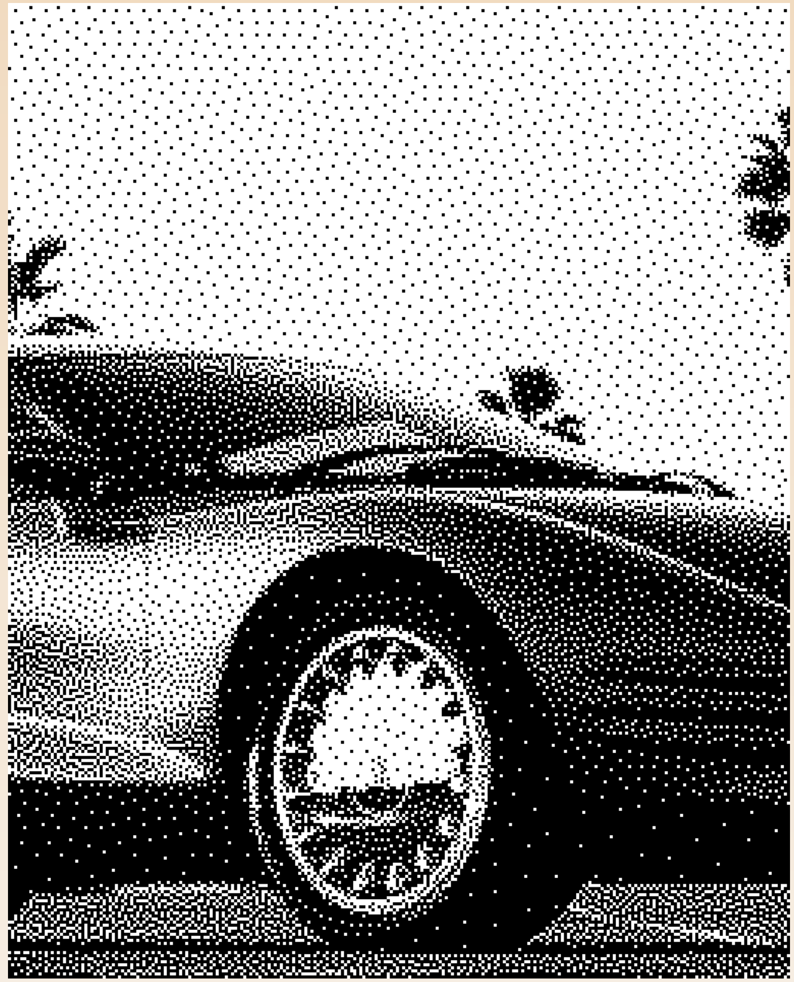
Optimal weights and thresholds



Floyd-Steinberg vs TDED

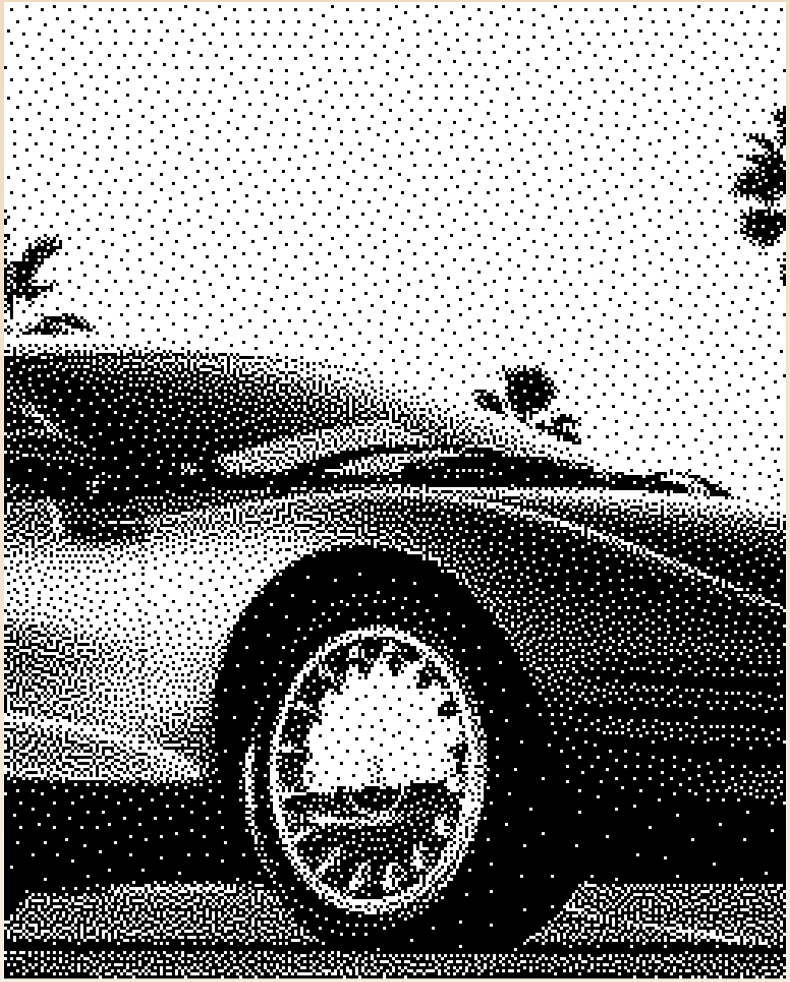


Floyd-Steinberg

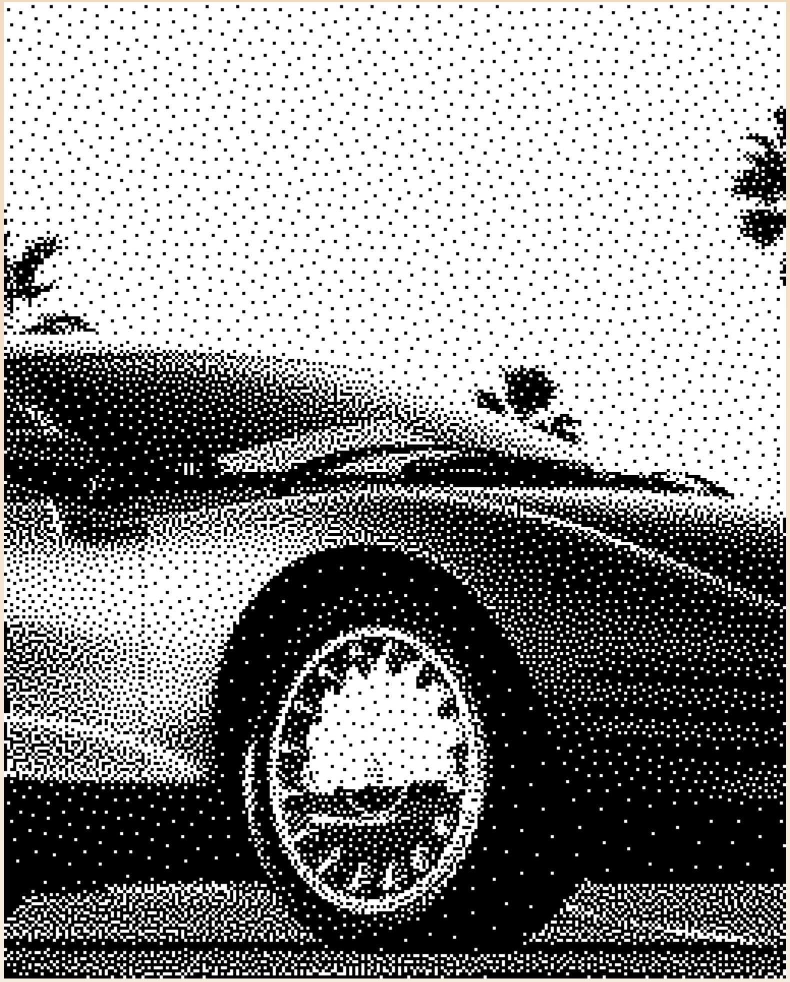


TDED

TDED vs DBS

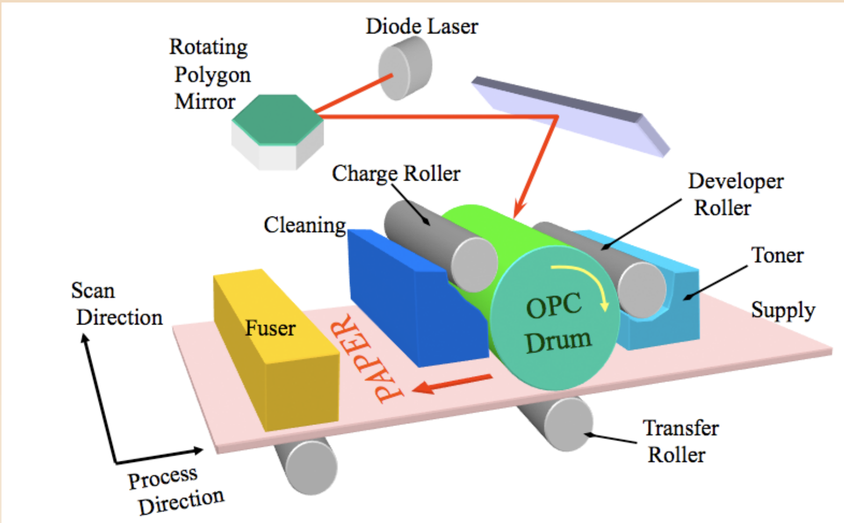


TDED



DBS

Marking engine technologies: laser electrophotographic



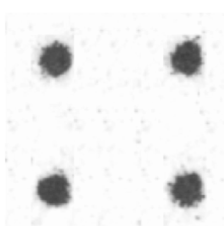
Architecture of laser electrophotographic printer

Typical low-end laser electrophotographic printer: HP LaserJet M252dw $249.99 List

Instability of electrophotographic process

Minimum Variance Pattern

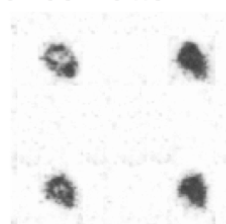
0	1	1
1	1	1
0	1	1



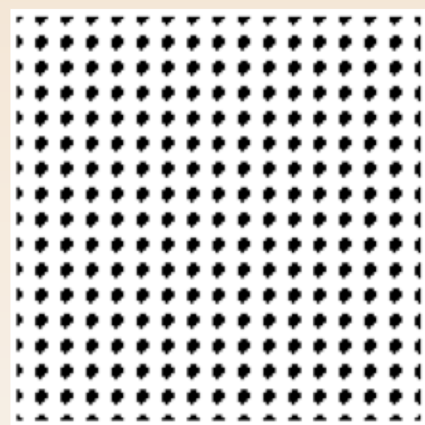
Avg. = 0.76, Std. Dev. = 0.089

Maximum Variance Pattern

0	1	0
1	1	1
0	0	1



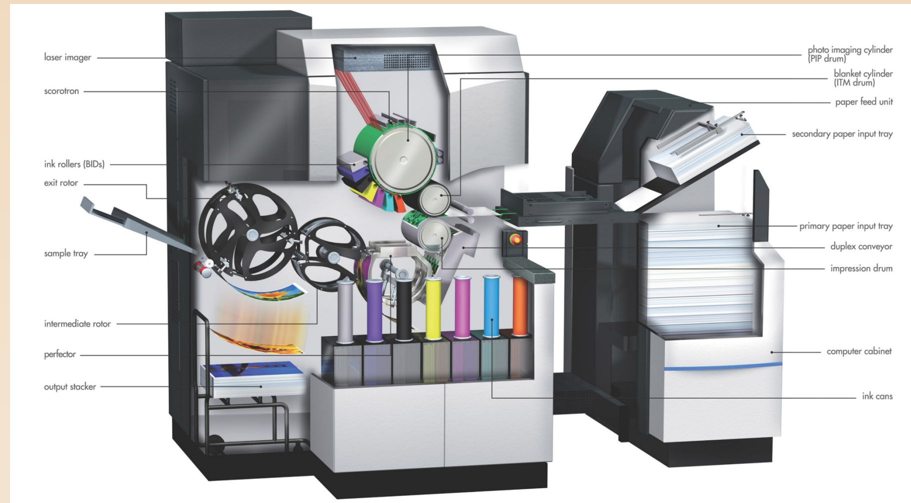
Avg. = 0.58, Std. Dev. = 0.480



Periodic, clustered-dot halftone textures are generally preferred for electrophotographic printers

Student: F. Baqai

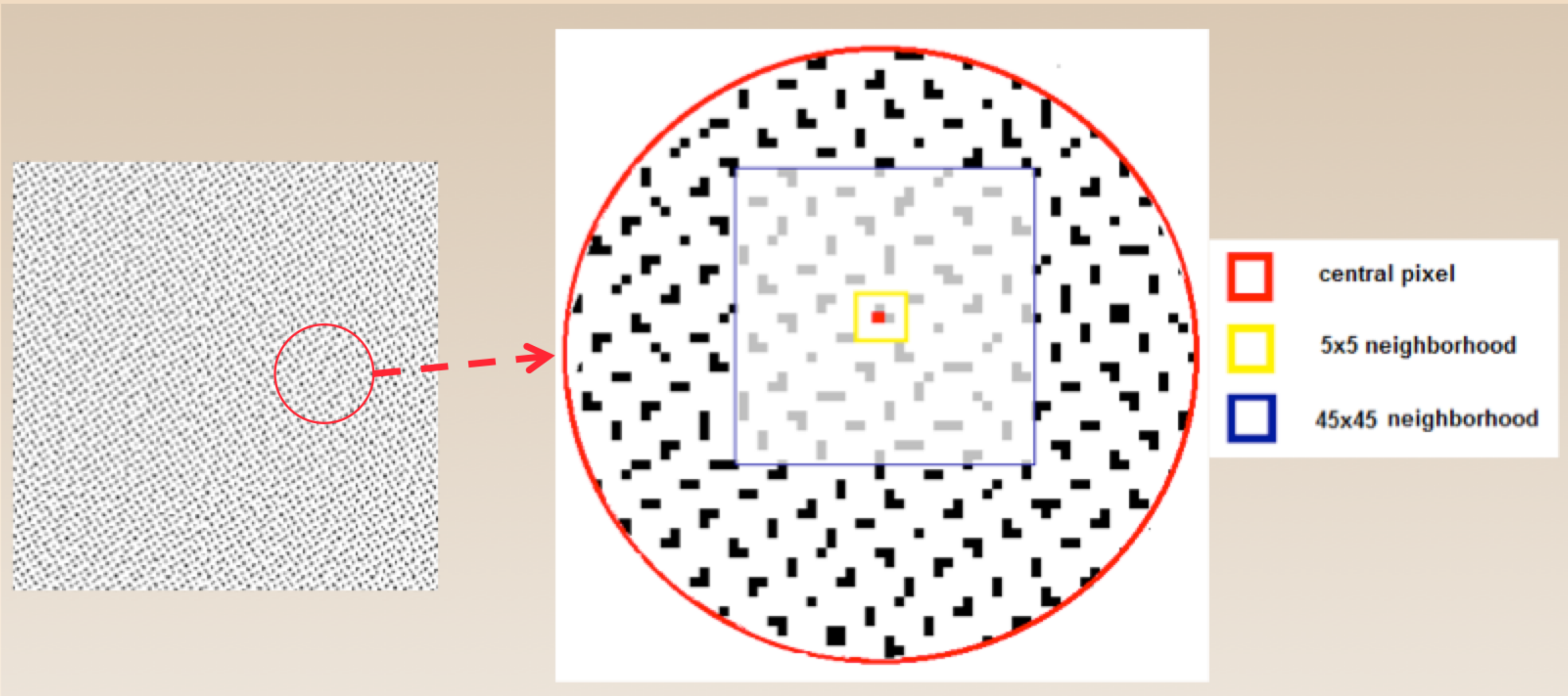
Commercial/industrial scale electrophotographic printing



HP Indigo Press 3050 2,000 4-color sheets/hr.

HP Indigo Press 30000
4600 3-color sheets/hr.

Linear Regression
Predicting Printed Absorptance From a Digital Halftone: the Black-Box Model*

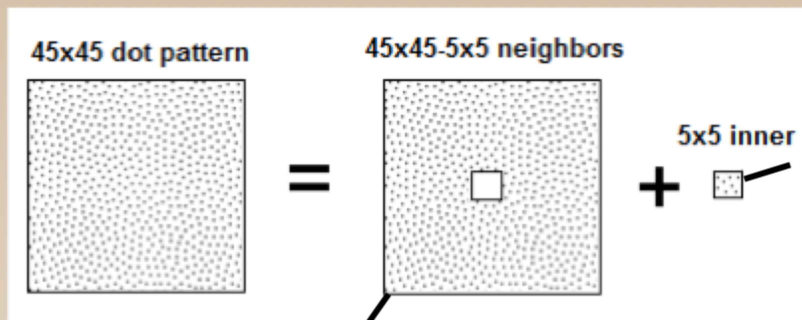


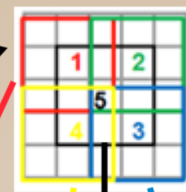
Y. Ju, T. Kashti, T. Frank, D. Kella, D. Shaked, M. Fischer, R. Ulichney, and J. P. Allebach, "Black-Box Models for Laser Electrophotographic Printers – Recent Progress," *Proceedings NIP29: IS&T's 29th International Conference on Digital Printing Technologies*, Seattle, WA, 29 September – 3 October 2013

*Research supported by HP, Inc.

Structure of the Black-Box Model

Predict the central pixel absorptance



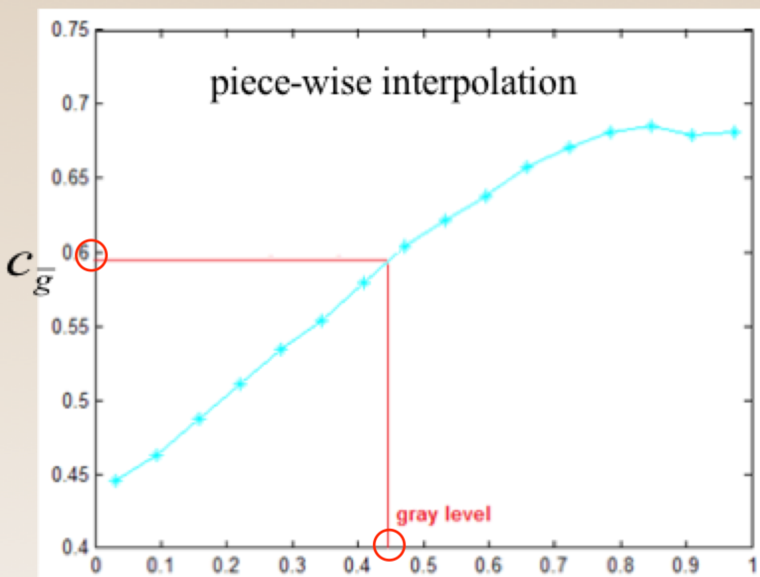


index:

$$\chi_j = \sum_{i=0}^{8} 2^i \cdot a_{i,j}$$

$$\bar{g} = \frac{\sum_{i \in \Omega} a_i}{45\times 45 - 5\times 5} \qquad \Omega : 45\times 45 - 5\times 5 \text{ neighbors}$$

a_i : digital halftone binary pixel value



χ_1	$w^2_{\chi_2}$
0	-0.076
1	-0.072
…	…
102	0.024
…	…
511	0.134

χ_2	$w^2_{\chi_2}$
0	-0.076
1	-0.071
…	…
201	0.016
…	…
511	0.172

χ_3	$w^3_{\chi_3}$
0	-0.066
1	-0.067
…	…
51	-0.004
…	…
511	0.434

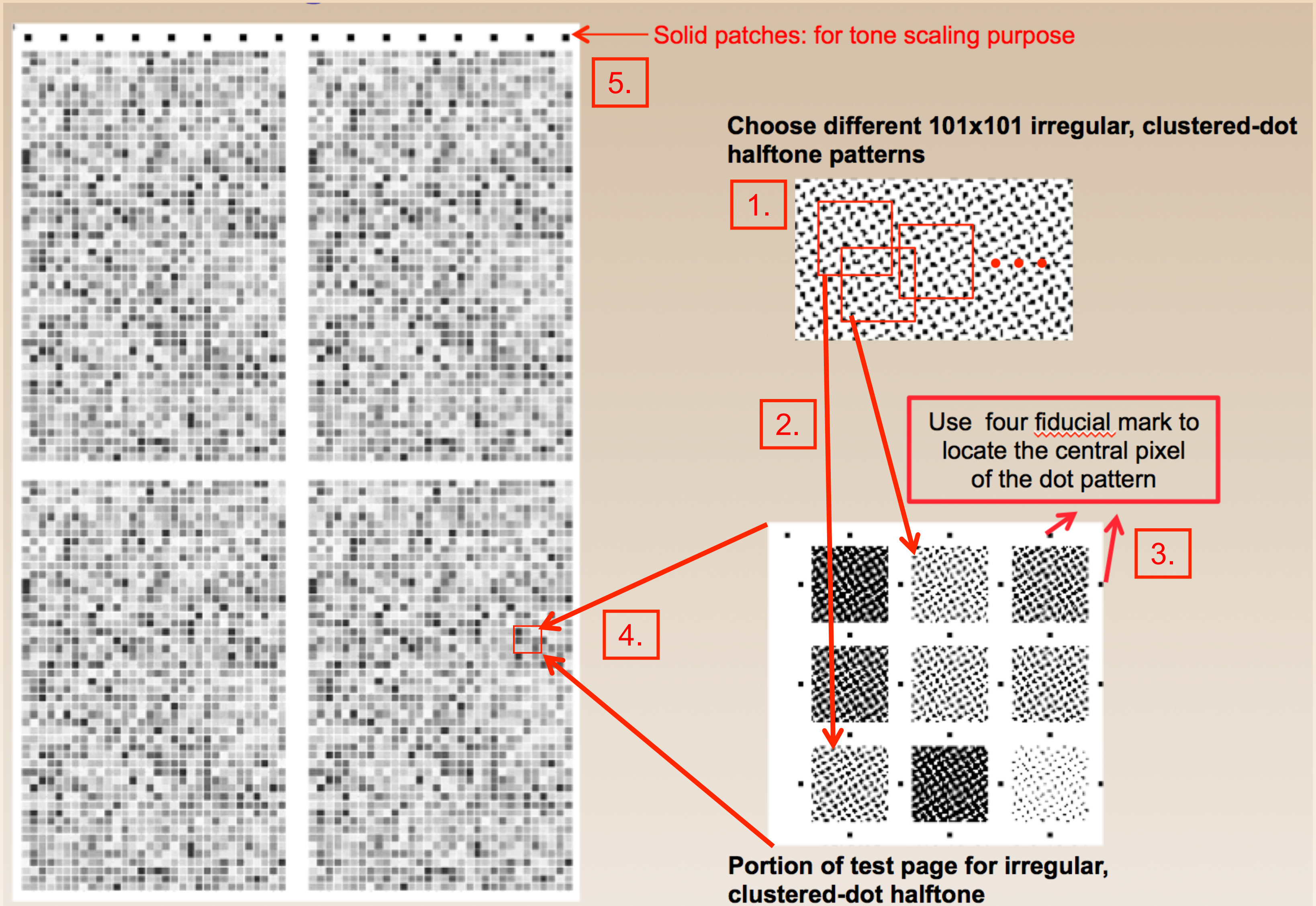
χ_4	$w^4_{\chi_4}$
0	-0.070
1	-0.062
…	…
12	0.001
…	…
511	0.172

χ_5	$w^5_{\chi_5}$
0	-0.154
1	-0.129
…	…
178	0.049
…	…
511	0.154

$$\hat{g}(\chi_1, \ldots, \chi_5; \bar{g}) = w^1_{\chi_1} + w^2_{\chi_2} + w^3_{\chi_3} + w^4_{\chi_4} + w^5_{\chi_5} + c_{\bar{g}}$$

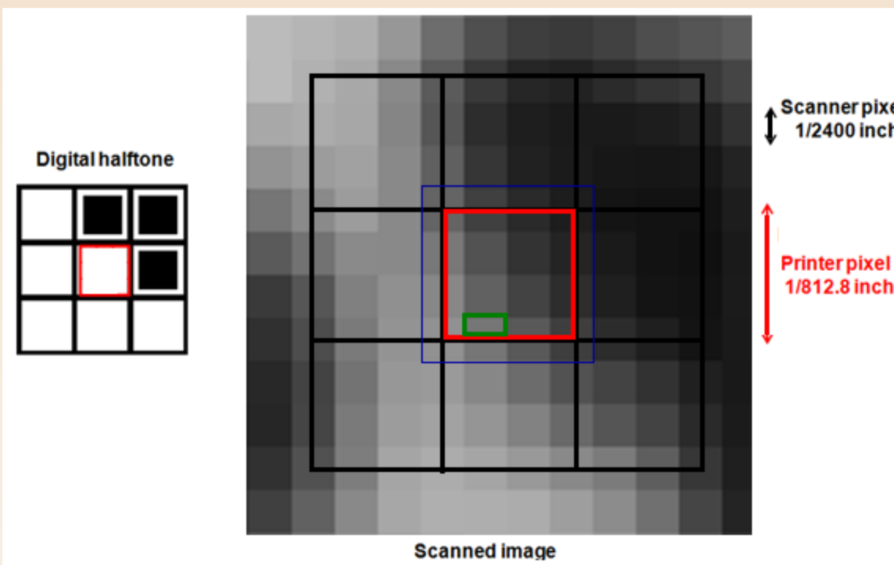
PURDUE
UNIVERSITY

How Do We Train the Model?



Scanned Image Analysis

Calibrated scanned image → Locate the centroid of each fiducial mark → Locate all pixels that have 45x45 surrounding → Estimate absorptance for all pixels within the region of interest → Statistics data for black box models



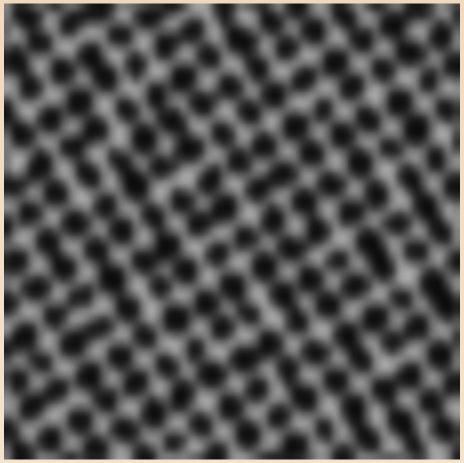
1 printer pixel = $\frac{2400}{812.8} \times \frac{2400}{812.8}$ scanner pixels

weight for $\square$ pixel = $\dfrac{0.5}{\frac{2400}{812.8} \times \frac{2400}{812.8}}$

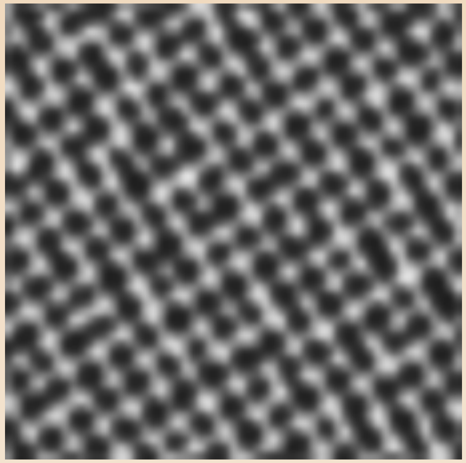
$$\tilde{g}[m,n] = \sum_{[k,l] \in \Omega_{m,n}} \omega_{m,n}[k,l]\, s[k,l]$$

Experimental Results – Sample Images

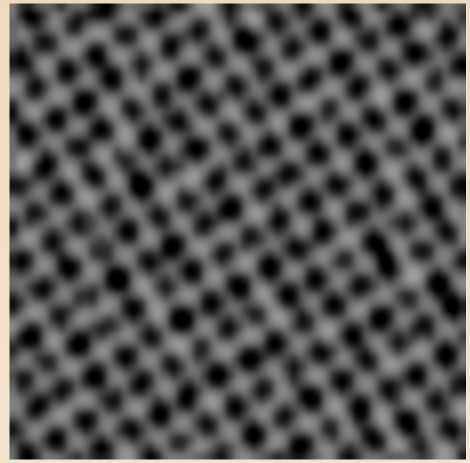
Scanned image



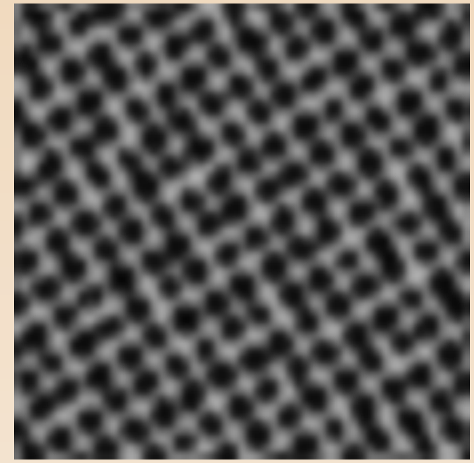
ULM5x5 prediction



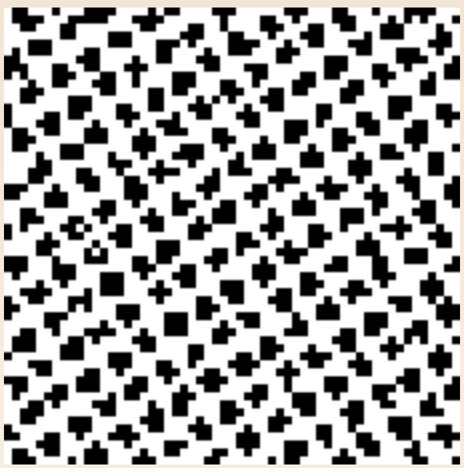
M45x45 c2a prediction



M45x45 c3b prediction

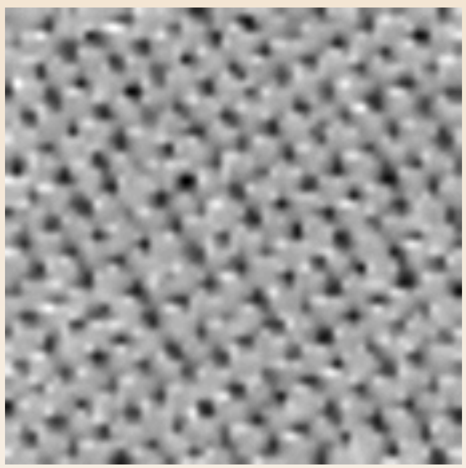


Digital

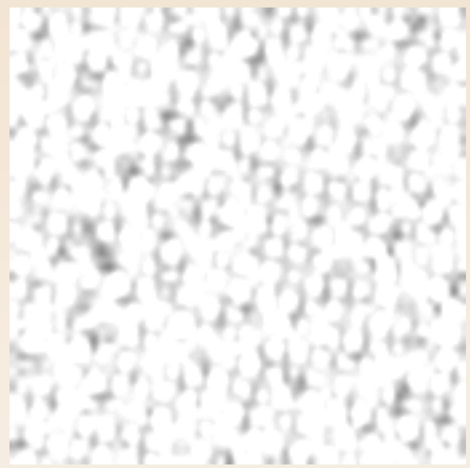


Gray level 96/255

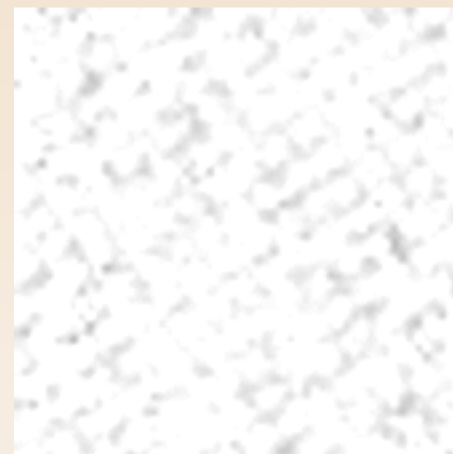
ULM5x5 error image*



M45x45 c2a error image

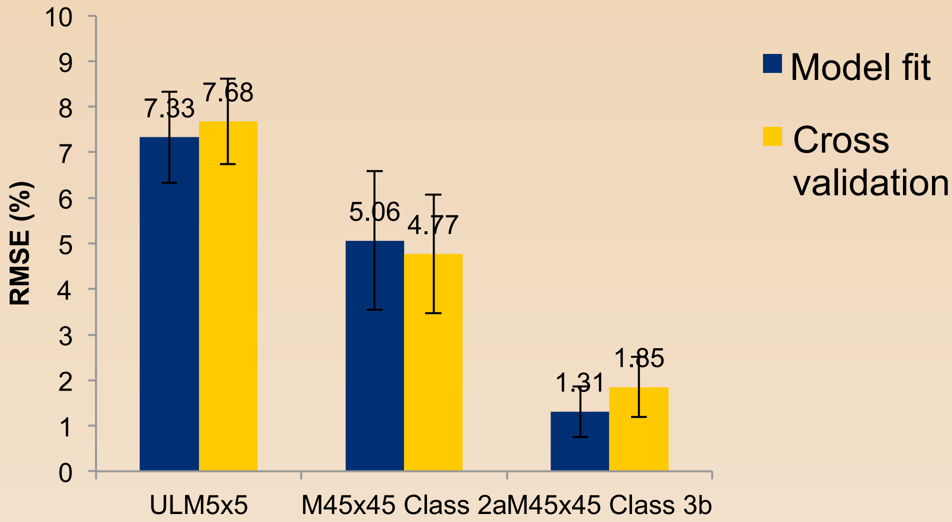


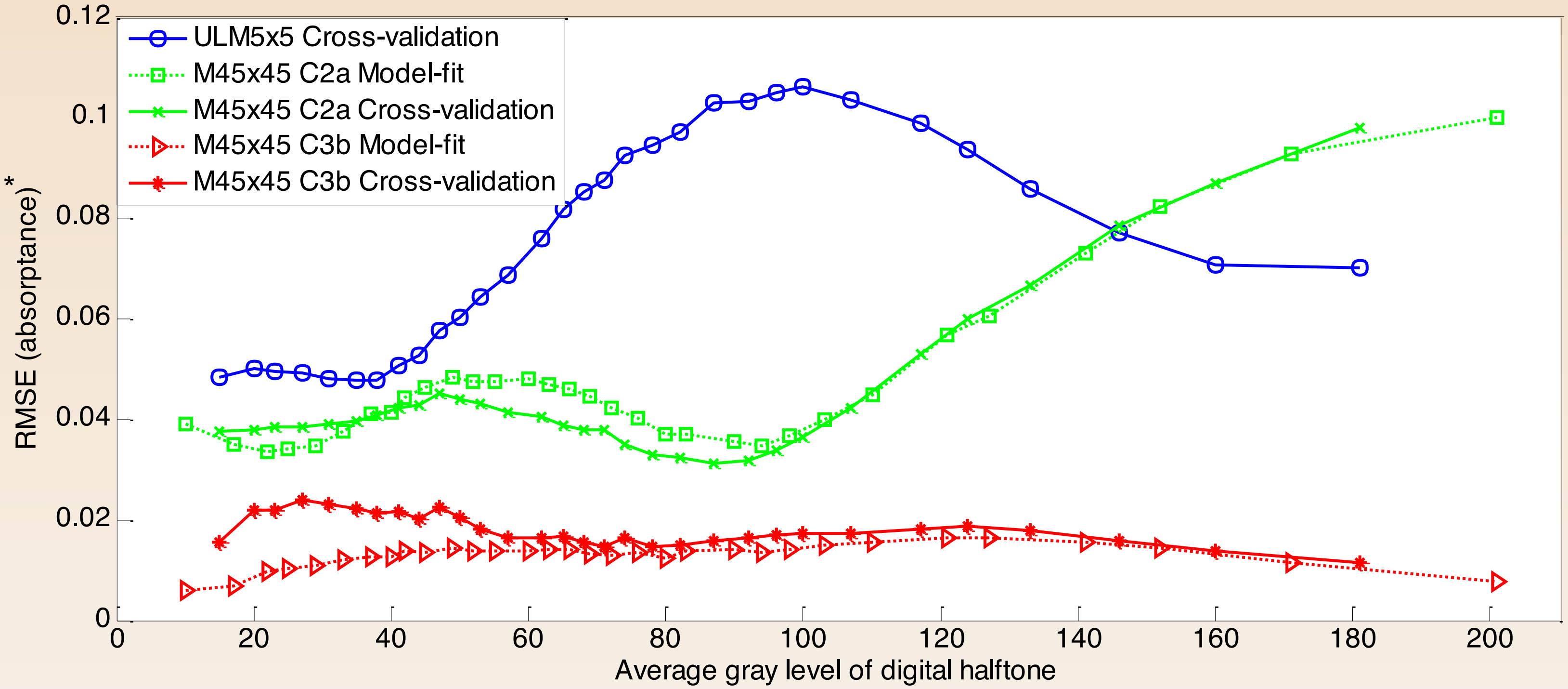
M45x45 c3b error image



*All error images are scaled identically with white denoting low error and black denoting high error.

Experimental Results – Error Statistics



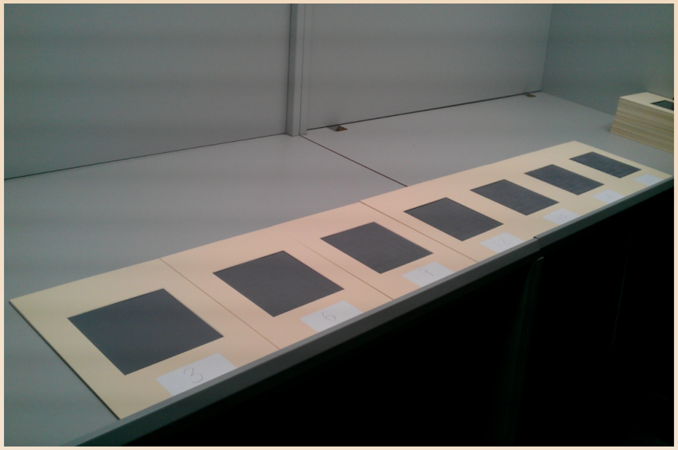


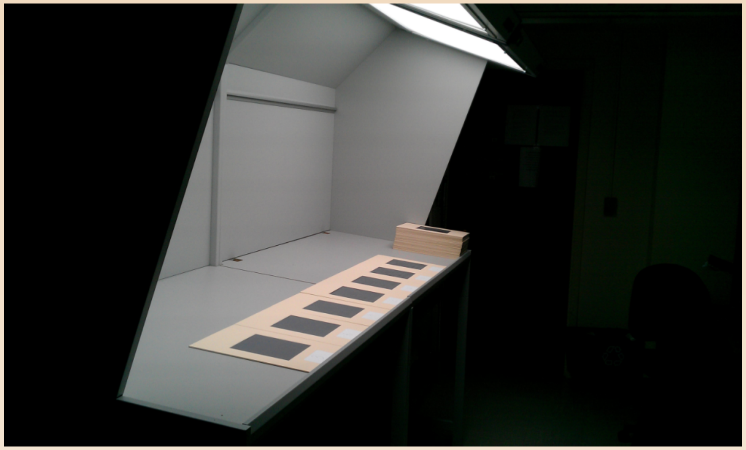
*Absorptance units are on a scale of 0 (white) to 1 (black)

Synopsis

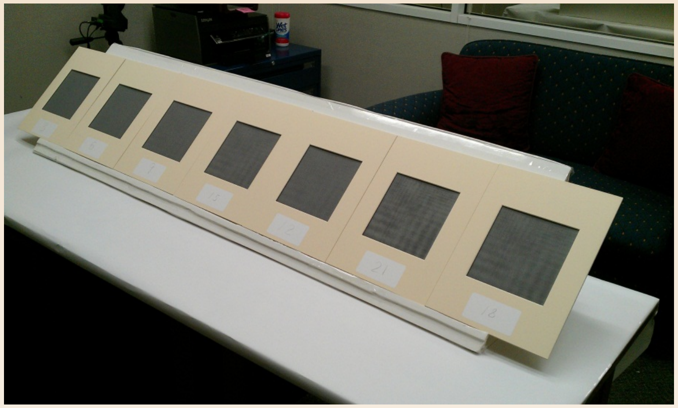
- K Nearest Neighbor classification applied to printer forensics
- Extension of K Means to Scalar Sequential Quantization
- Optimal tree-structured piece-wise linear filter for image scaling
- Training-based methods for digital haftoning
- Black-box model for print prediction based on training and linear regression
- Print macrouniformity prediction (Method 1)
- Print macrouniformity prediction (Method 2)
- Fashion photograph aesthetic quality predictor based on SVM and CNN
- Facial landmark detection using CNN
- Logo identification using CNN
- Text field category classification via natural language processing

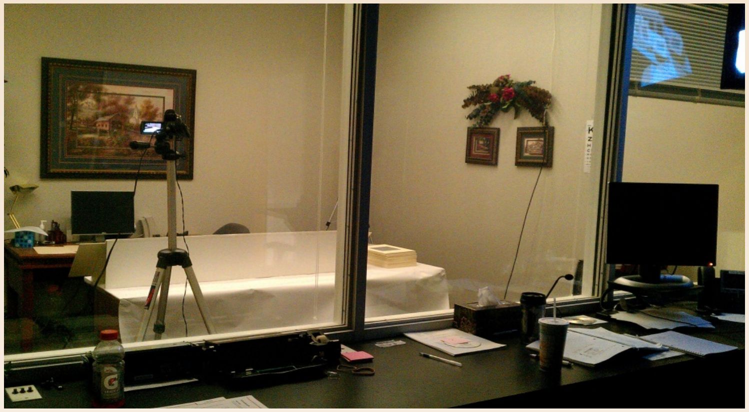
Linear Regression and Support Vector Machine Assessment of Large Area Nonuniformity by Image Quality Ruler Method*





Experimental set-up at Purdue University



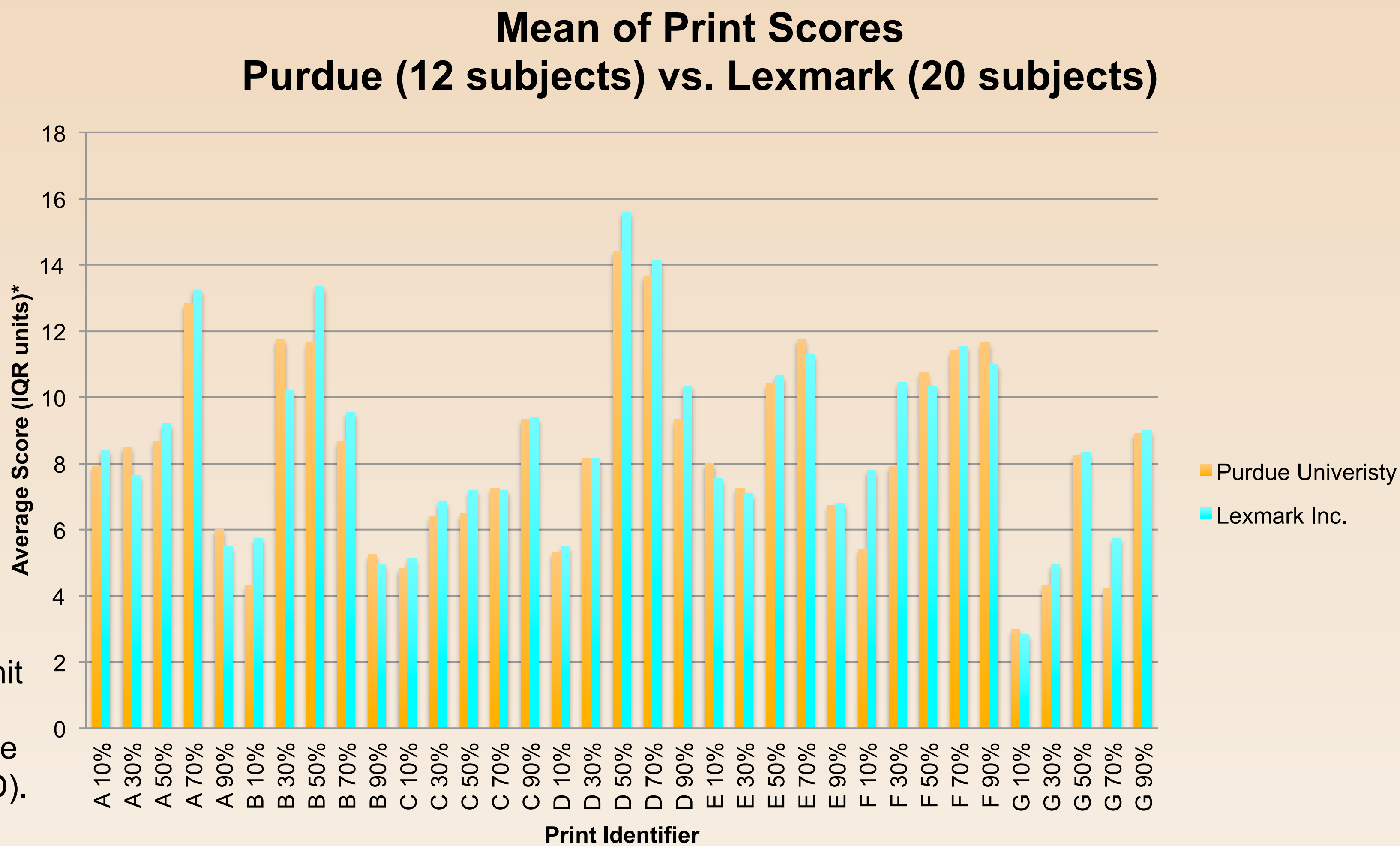


Experimental set-up at Lexmark site

W. Wang, G. Overall, T. Riggs, R. Silveston-Keith, J. Whitney, G. T. C. Chiu, and J. P. Allebach, "Figure of Merit for Macrouniformity Based on Image Quality Ruler Evaluation and Machine Learning Framework," *Image Quality and System Performance X,* SPIE Vol. 8653, P. D. Burns and S. Triantaphillidou, Eds. San Francisco, CA, 3-7 February 2013.

Research supported by Lexmark

Results from Image Quality Ruler Experiment for Assessment of Macro-Uniformity



*Each IQR unit represents 1 just-noticeable different (JND).

Lower scores correspond to higher quality.

Mean difference between Purdue and Lexmark scores is 0.66 and the correlation is 0.95.

PURDUE UNIVERSITY

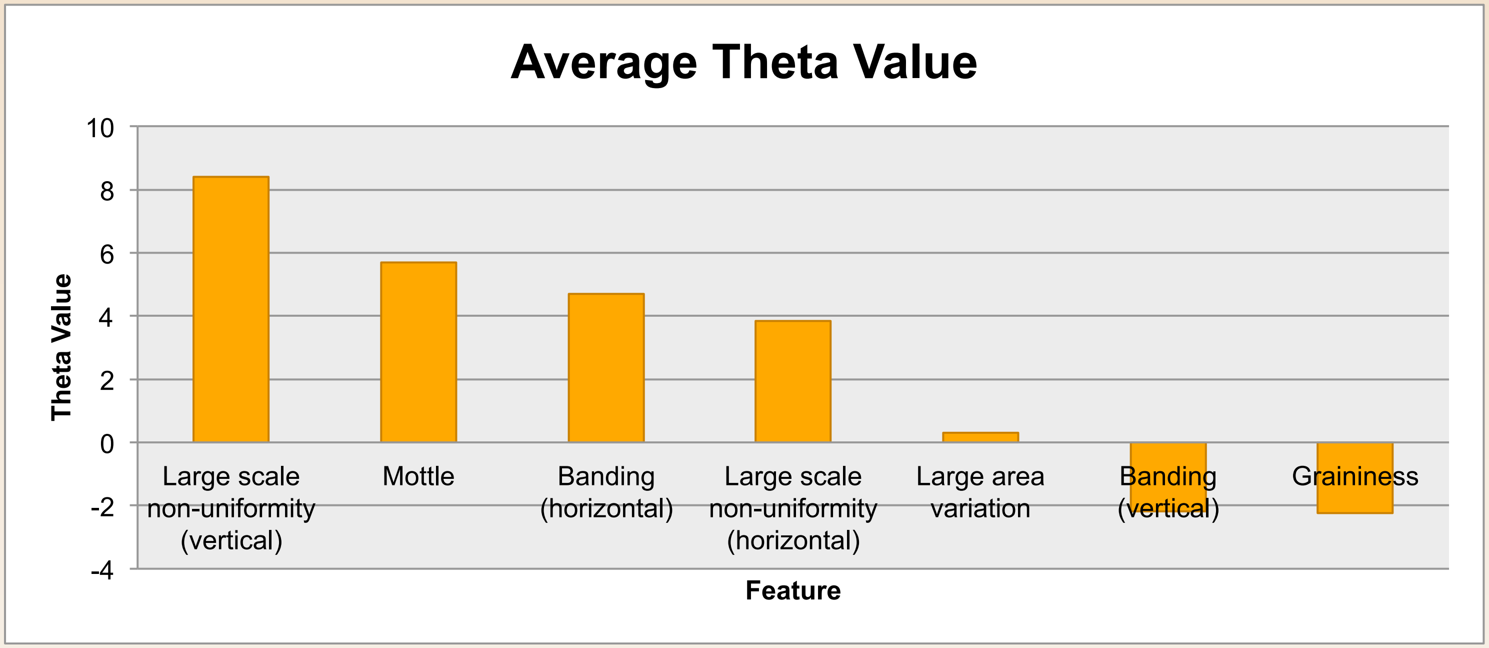
Prediction of Scores Assigned by Human Observers: Macro-Uniformity Features

- Graininess: 2-dimensional, grainy texture.
- Mottle: 2-dimensional, random lightness variations.
- Large area variation: 2-dimensional, random lightness variations, spatial region is larger than mottle.
- Jitter (horizontal and vertical): 1-dimensional, isolated lightness variations.
- Large-scale non-uniformity (horizontal and vertical): 1-dimensional, periodic lightness variations.
- The algorithms that we used are largely inspired by ISO image quality standards.*

*Document B123: NP 13660 office equipment measurement of image quality attributes for hardcopy output: Binary monochrome text and graphic images, ISO/IEC.

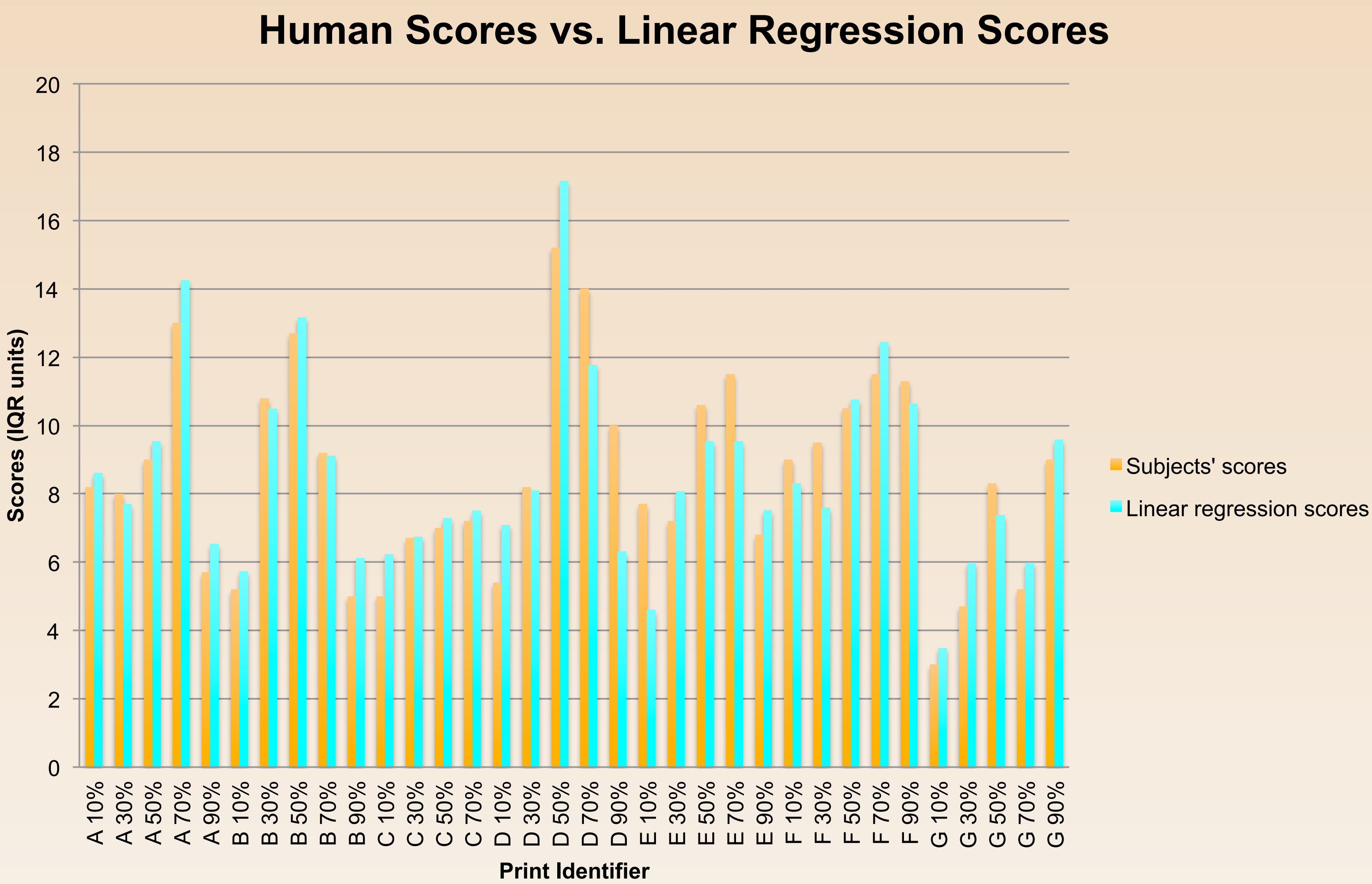
Prediction of Macro-Uniformity Scores by Linear Regression

- Predicted Rating = $\theta_0 + \theta_1 \times f_1 + \theta_2 \times f_2 + \ldots$
- Training error
  - » Mean absolute error is 0.80, standard deviation of error is 0.64
- Testing error
  - » Mean absolute error is 0.98, standard deviation of error is 0.83



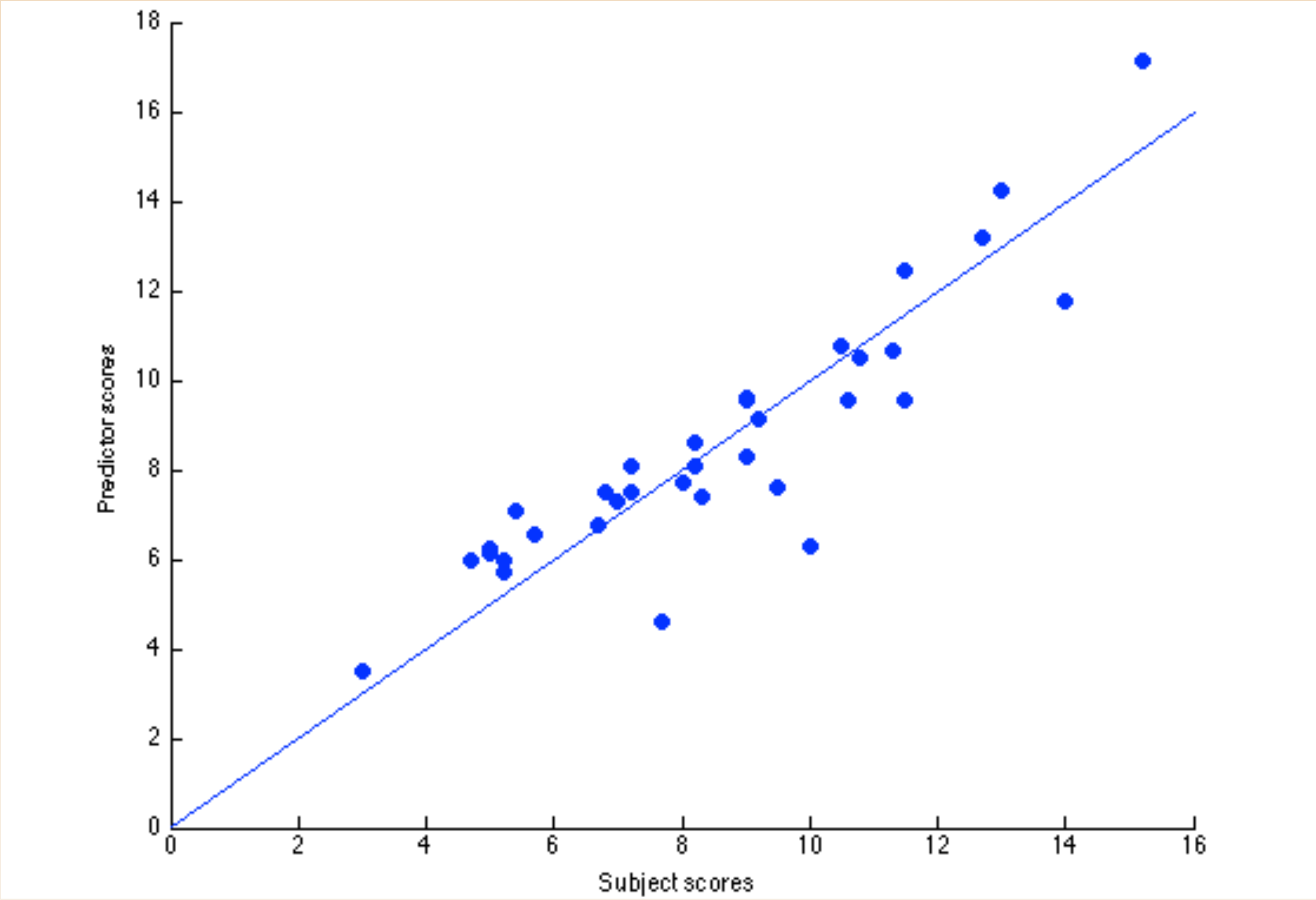
PURDUE UNIVERSITY

Accuracy of Macro-Uniformity Predictor as a Function of Print Sample



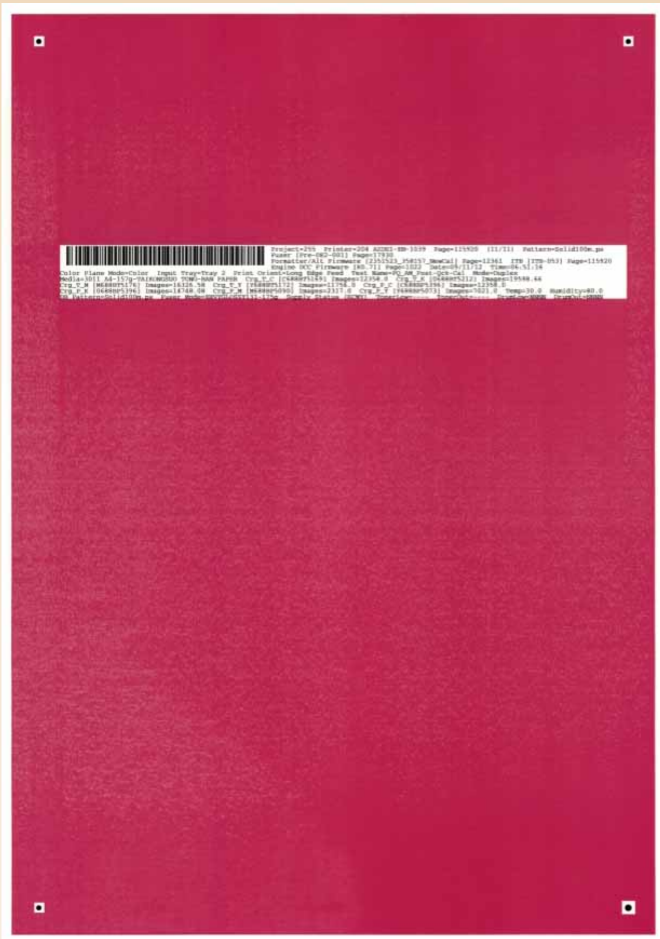
PURDUE UNIVERSITY

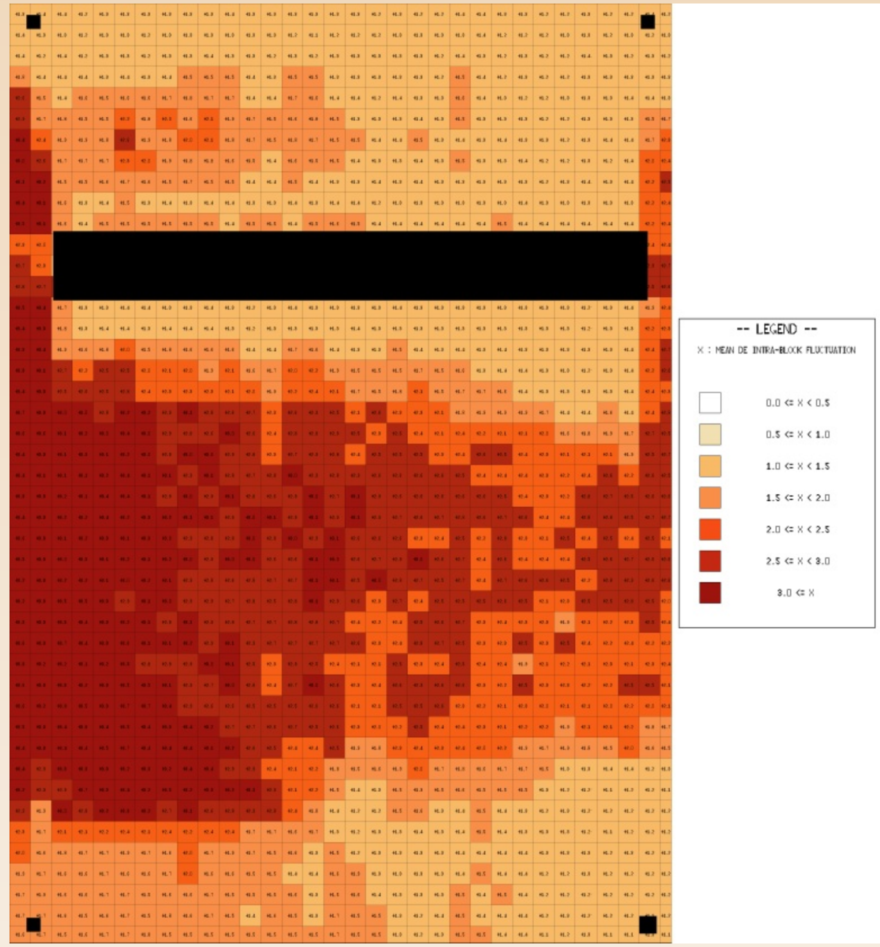
Accuracy of Macro-Uniformity Predictor in Terms of Scatter Plot



The correlation between Linear Regression Predicted Scores and Subjects' Scores is 0.90

Support Vector Machine Assessment of Local Nonuniformity*





M. Q. Nguyen, S. Astling, R. Jessome, E. Maggard, T. Nelson, M. Q. Shaw, and J. P. Allebach, "Perceptual Metrics and Visualization Tools for Evaluation of Page Uniformity," *Image Quality and System Performance XI,* SPIE Vol. 9016, S. Triantaphillidou and M.-C. Larabi, Eds. San Francisco, CA, 3-5 February 2014.

M. Q. Nguyen and J. P. Allebach, "Controlling Misses and False Alarms in a Machine Learning Framework," *Image Quality and System Performance XII,* SPIE Vol. 9396, M.-C. Larabi and S. Triantaphillidou, Eds. San Francisco, CA, 8-12 February 2015.

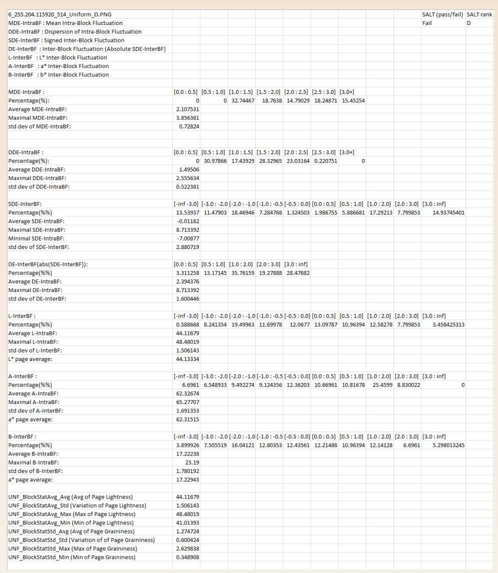
*Research supported by HP, Inc.

PURDUE UNIVERSITY

Prediction of Non-Uniformity Grades Assigned by an Expert Human Observer: Data Set and Features

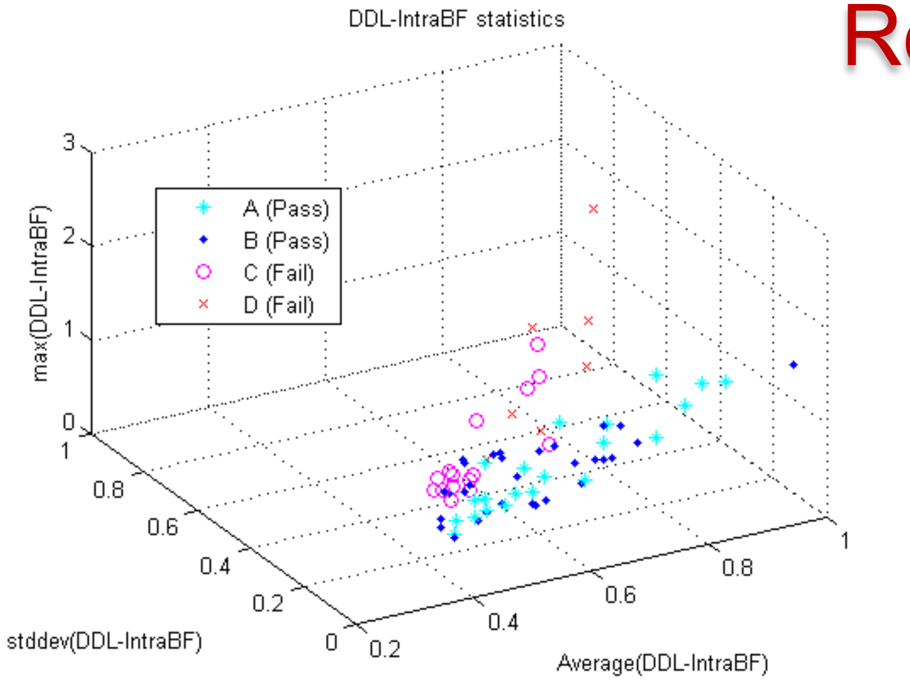
Total	251	Print Quality	P/F
Rank A	24	good	pass
Rank B	136	fairly good	pass
Rank C	66	bad	fail
Rank D	25	very bad	fail

- Each test page includes 40 statistics from 8 features (histogram, min, max, mean, stddev)

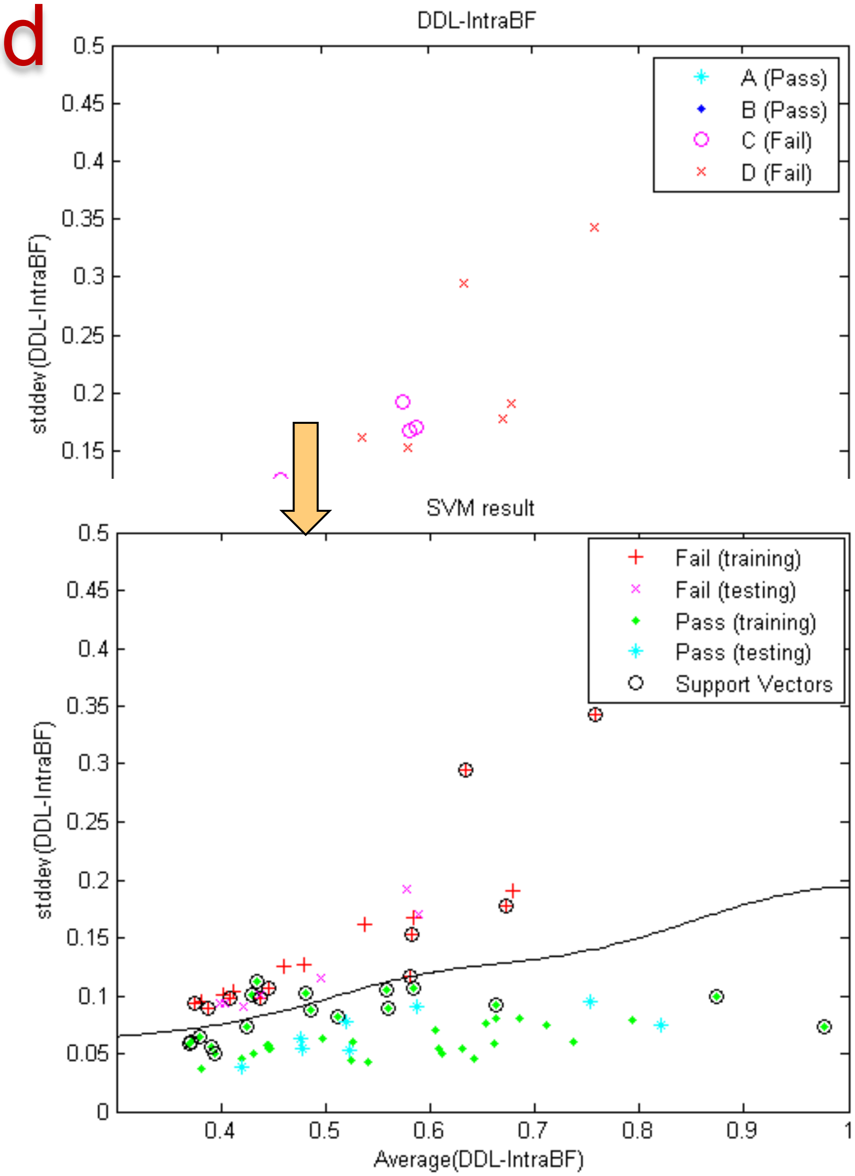


PURDUE UNIVERSITY

Use of Support Vector Machine (SVM) to Predict Non-Uniformity Grades Assigned by Expert Observer



Red



- For each test page, there are 40 statistics from 8 features (histogram, min, max, mean, stddev)
- For SVM, use DDL-IntraBF and SDE-InterBF (Gaussian radial basis, stddev = 1)
- Perform 5-fold cross validation.

Performance of SVM in Predicting Non-Uniformity Grades Assigned by Expert Observer

Table 1. SVM prediction results for 3 types of test pages

(a) Cyan with 83.3% correct classification.

Cyan (90 test pages)	SVM-based prediction	
Visual grade		Pass	Fail
	Pass	49	4
	Fail	11	26

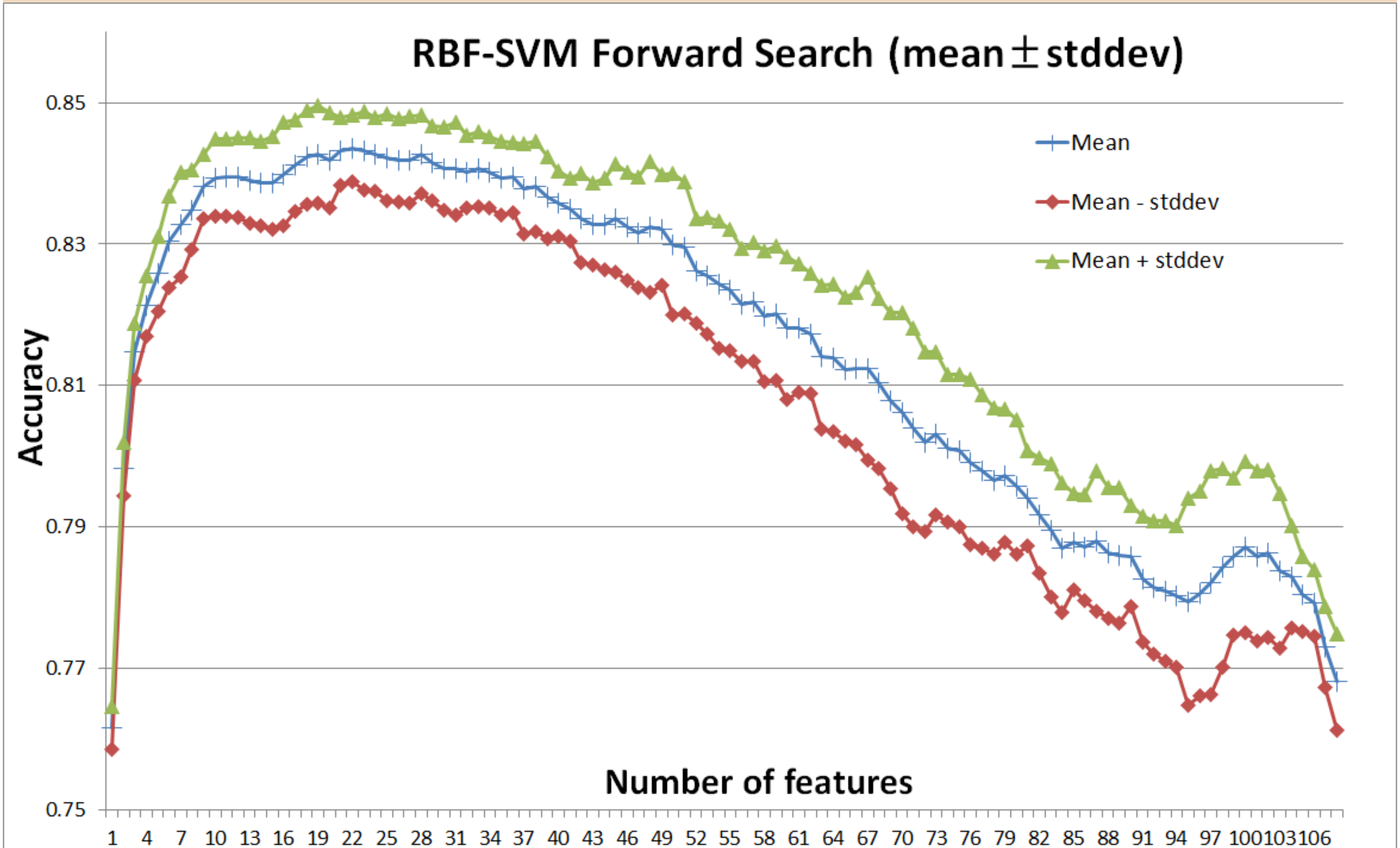
(b) Magenta with 85.2% correct classification.

Magenta (88 test pages)	SVM-based prediction	
Visual grade		Pass	Fail
	Pass	49	7
	Fail	6	26

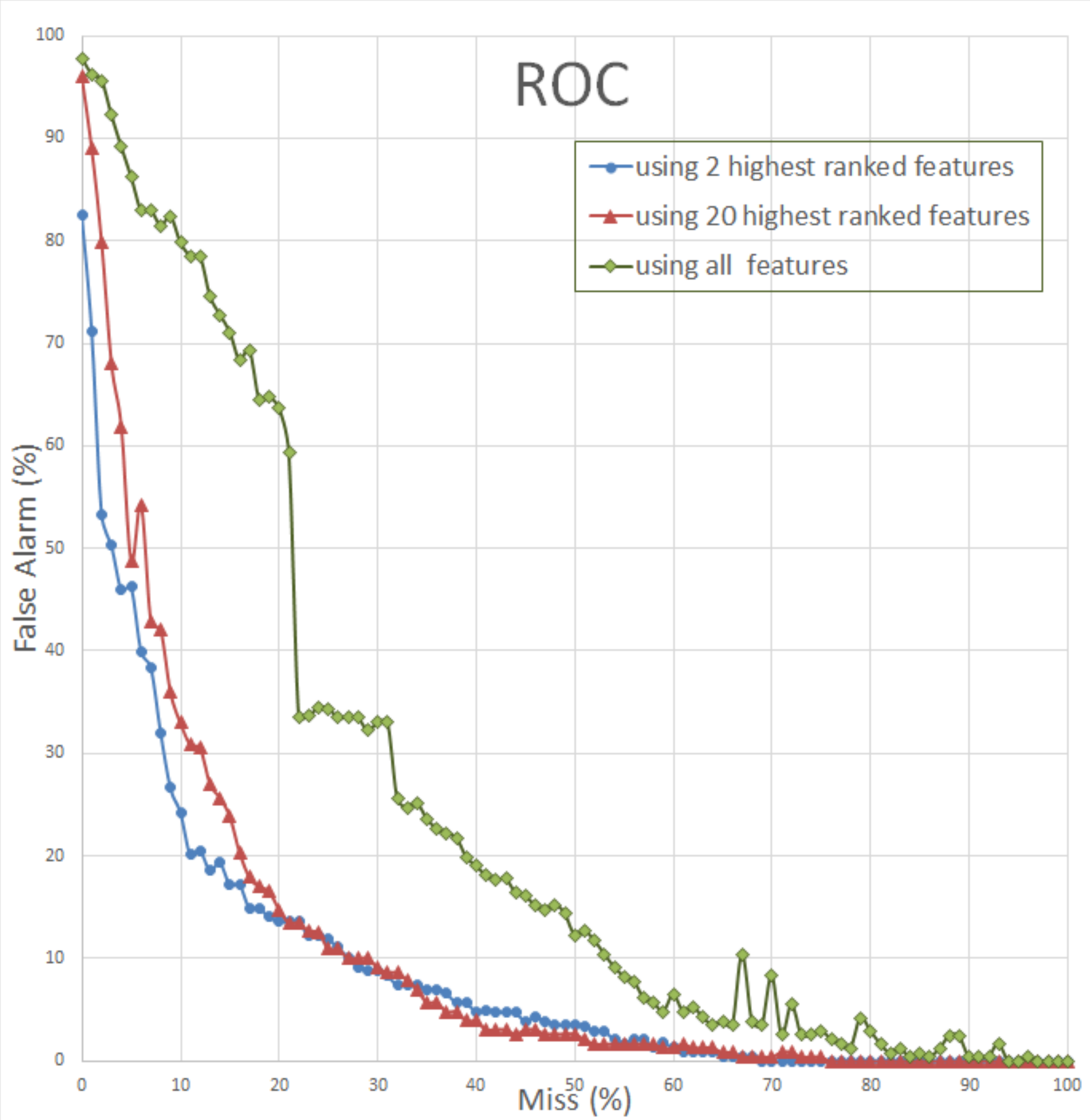
(c) Red with 90.4% correct classification.

Red (73 test pages)	SVM-based prediction	
Visual grade		Pass	Fail
	Pass	45	6
	Fail	1	21

Refinement of Feature Set by Forward Search



Controlling False Alarms vs. Misses



Synopsis

- K Nearest Neighbor classification applied to printer forensics
- Extension of K Means to Scalar Sequential Quantization
- Optimal tree-structured piece-wise linear filter for image scaling
- Training-based methods for digital haftoning
- Black-box model for print prediction based on training and linear regression
- Print macrouniformity prediction (Method 1)
- Print macrouniformity prediction (Method 2)
- Fashion photograph aesthetic quality predictor based on SVM and CNN
- Facial landmark detection using CNN
- Logo identification using CNN
- Text field category classification via natural language processing

Support Vector Machine and Convolutional Neural Network Fashion Photograph Aesthetic Quality Predictor*

- Goal is to develop a method to automatically generate aesthetic quality scores for photos.
  - » Focus on customer-uploaded fashion item photos on customer-to-customer (C2C) fashion shopping website. Mostly taken by amateur photographers.
  - » When customers upload item photos, we can give them feedback on the aesthetic quality. If the quality is not satisfactory, we may suggest customers retaking photos.
  - » Our sponsor can use the predictor to decide which closet is highlighted.

M. Chen and J. P. Allebach, "Aesthetic Quality Inference for Online Fashion Shopping," *Imaging and Multimedia Analytics in a Web and Mobile World 2014,* SPIE Vol. 9027, Q. Lin, J. P. Allebach, and Z. Fan, Eds. San Francisco, CA, 3-4 February 2014.

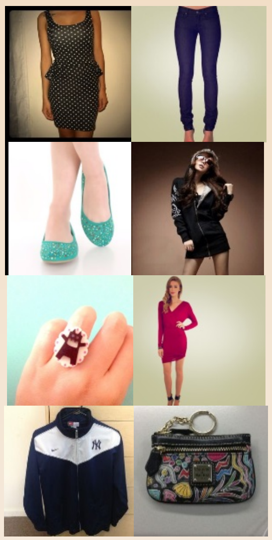
J. Wang, "Three Problems in Image Analysis and Rendering: Aesthetic Evaluation of Fashion Photos, Local Defect Detection, and Semantically-Based 2.5D Printing," Ph.D. Dissertation, Purdue University , West Lafayette, IN, May 2016.

*Research supported by Poshmark, Inc.

POSHMARK

Framework for aesthetic quality prediction

Training/ Testing photos



Ground Truth Score

Conduct psychophysical experiments :
Ask women participants to rate photos from a fashion website on scale from 1-10

Learning/ Inference

Support Vector Regression

Feature Extraction

Low Level Features:
Sharpness
Colorfulness
Lightness
Contrast

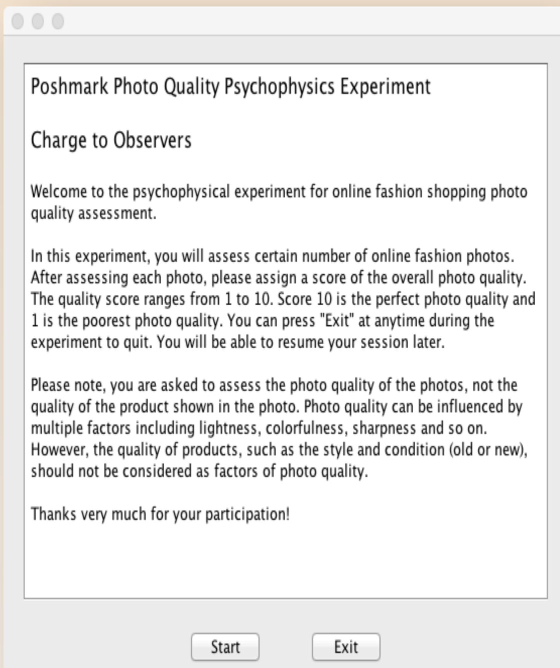
Salient Object Detection:
Salient region area
Salient region number
Subject to background difference

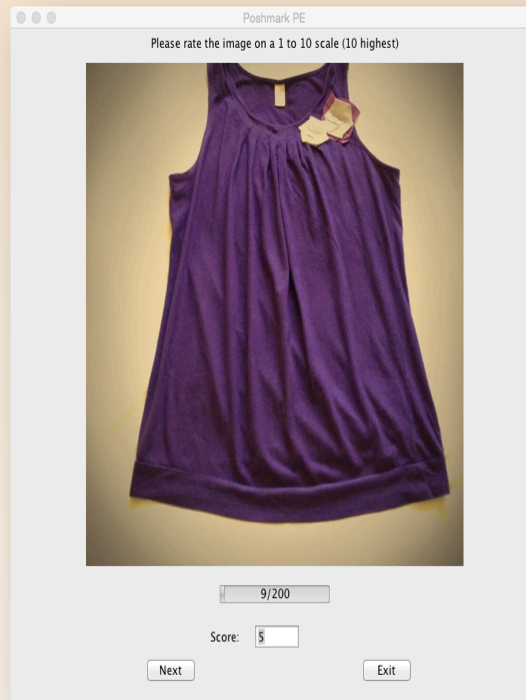
Metadata: Categories of items

Color Harmony
Hue Count
Modified Rule of Thirds
GIST

Ground Truth Collection

- We collected a dataset of 734 photos from our sponsor (www.poshmark.com).
  - » We built a GUI, and asked experiment participants to input the aesthetic quality score for each photo.
  - » The rating is based on a 1 to 10-point scale, where 1 denotes worst quality and 10 denotes best quality.





Example Feature: Colorfulness – Highest and Lowest 3 from Training and Testing Database

144.1

96.8

96.1

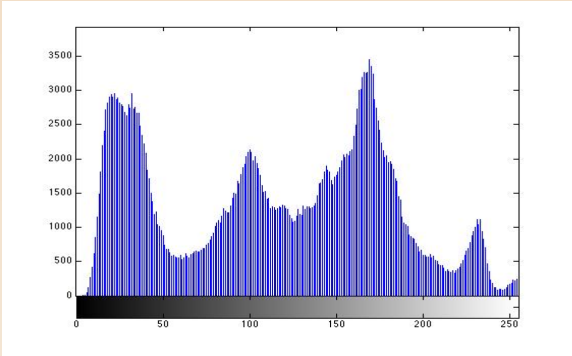
26.6

26.0

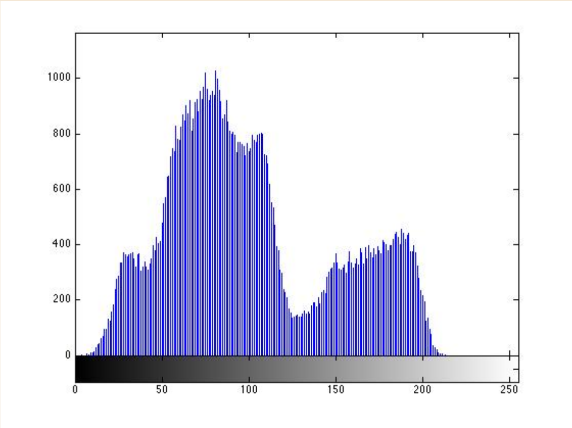
22.6

Example Feature: Contrast Metric

- The span of the histogram that contains the central 98% of gray levels of the image.

Contrast Score: 224

Contrast Score: 178

Example Feature: Saliency

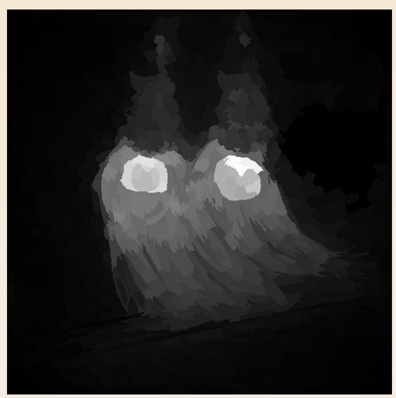
Original Image

Original Image

Original Image

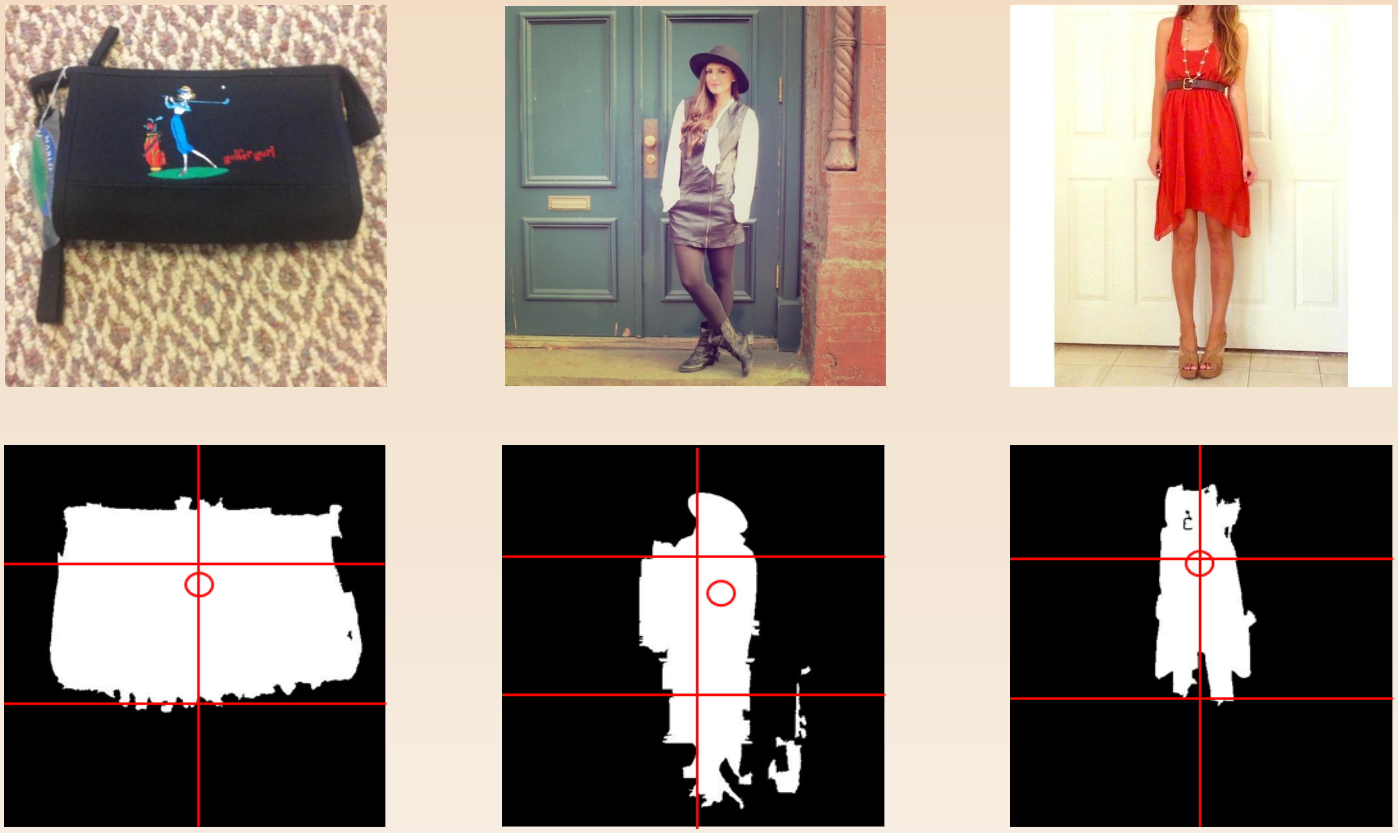
Saliency Map

Saliency Map



Saliency Map

Example Feature: Modified Rule of Thirds



32 80 3

Ground Truth and Predicted Aesthetic Scores
Examples of High and Low Quality Photos

Predicted Score: 7.9
Ground Truth Score: 8.6

Predicted Score: 7.6
Ground Truth Score: 8.8

Predicted Score: 9.6
Ground Truth Score: 9

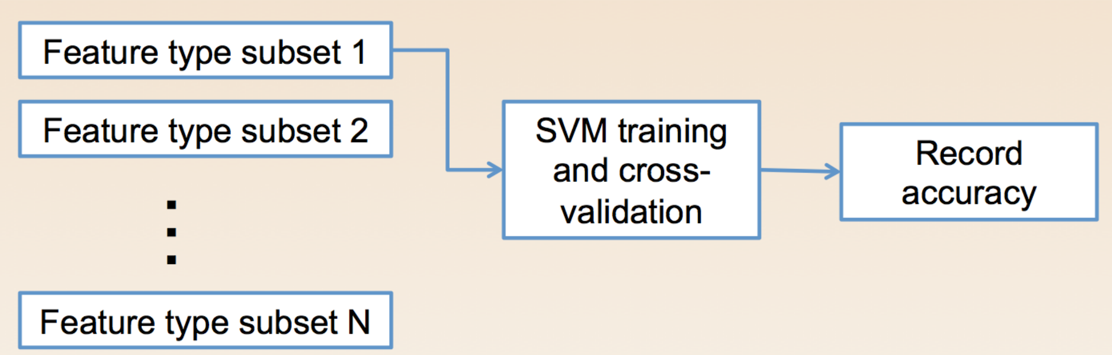
Predicted Score: 4.1
Ground Truth Score: 5.2

Predicted Score: 2.6
Ground Truth Score: 4.1

Predicted Score: 2.1
Ground Truth Score: 4.8

Optimal Training Feature Subset Selection

- Using a subset of all designed features in predictor training may yield better result.
  - Mainly because overfitting is alleviated.
- Adopt wrapper feature selection methodology*.
  - Evaluate a feature subset by assessing the cross-validation accuracy of the SVR predictor trained with this feature subset.
  - In the end, we choose the feature subset that yields highest cross-validation accuracy.



*Isabelle Guyon and André Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

Wrapper Feature Selection Procedure and Result

- Exhaustively searching over all possible feature subsets is computationally intractable.
  - In our case 2^{26} passes would be needed.
  - We adopt the *best-first* algorithm as our search strategy*.
- Feature subset with the 9 selected features shown in the table can train a most accurate predictor.
  - However, if we are able to collect more training data, more features should be included since larger training dataset can support a model with higher complexity.

Feature ID	Feature Name	Selected in 734-photo dataset
1	'Lightness'	x
2	'Colorfulness'	
3	'Contrast'	
4	'Average Saturation'	
5	'Average Hue'	
6	'Hue Count'	
7	'Color Harmony'	
8	'Number Of Salient Regions'	
9	'Aggregate Area Size Of All Salient Regions'	
10	'Subject-Background Lightness Difference'	
11	'Subject-Background Hue Difference'	
12	'Subject-Background Saturation Difference'	x
13	'Modified Rule Of Thirds'	x
14	'Wavelet Level 1 Sum Of Power'	
15	'Wavelet Level 2 Sum Of Power'	
16	'Wavelet Level 3 Sum Of Power'	
17	'Laplacian Level 1 Sum Of Power'	x
18	'Laplacian Level 2 Sum Of Power'	x
19	'Laplacian Level 3 Sum Of Power'	x
20	'CPBD Sharpness Metric'	x
21	'Wavelet Center Detail Strength Ratio'	
22	'Laplacian Center Detail Strength Ratio'	x
23	'Wavelet Edge Energy Bounding Box'	
24	'Laplacian Edge Energy Bounding Box'	
25	'Wavelet Sum Of Weighted Distance'	
26	'Laplacian Sum Of Weighted Distance'	x

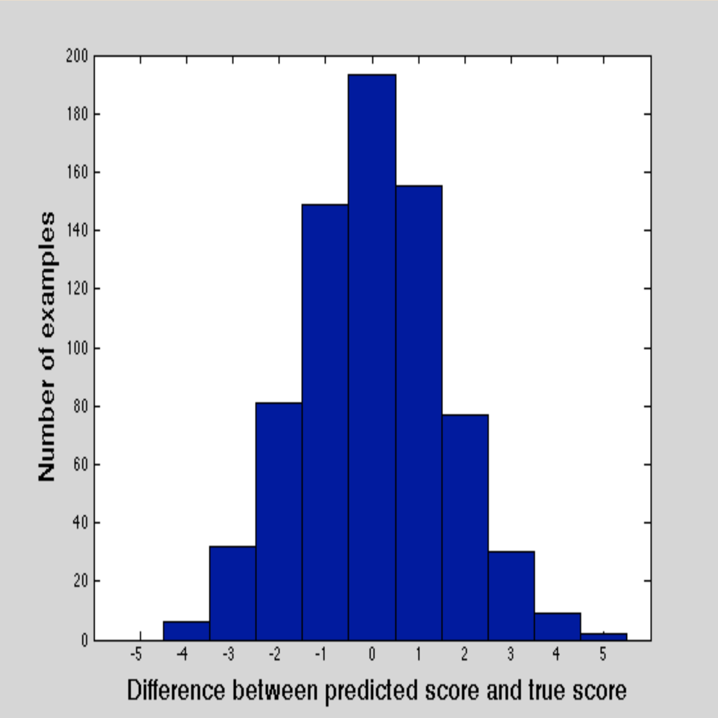
*Mark A Hall, *Correlation-based Feature Selection for Machine Learning*, Ph.D. thesis, The University of Waikato, 1999.

Predictor Training and Accuracy Analysis

- Database consists of 734 fashion shopping photos.
  - » Ground truth collected from psychophysical experiment.
- Trained a support vector regression predictor.
- Prediction accuracy analysis
  - » We conducted 10 repetitions (with different random partitions) of 10-fold cross-validation and calculated the average root mean squared error (RMSE) between the predicted aesthetic score and the ground truth.
  - » Using all the features, our regression predictor achieves an RMSE of 1.60 (score ranging from 1 to 10).
  - » With the optimal feature subset selected with wrapper feature selection, we further get an RMSE of 1.54.

Prediction Difference Histogram

- 49.32% of the examples have absolute differences between ground truth and predicted score smaller than 1.
- 79.97% of the examples have absolute differences smaller than 2.

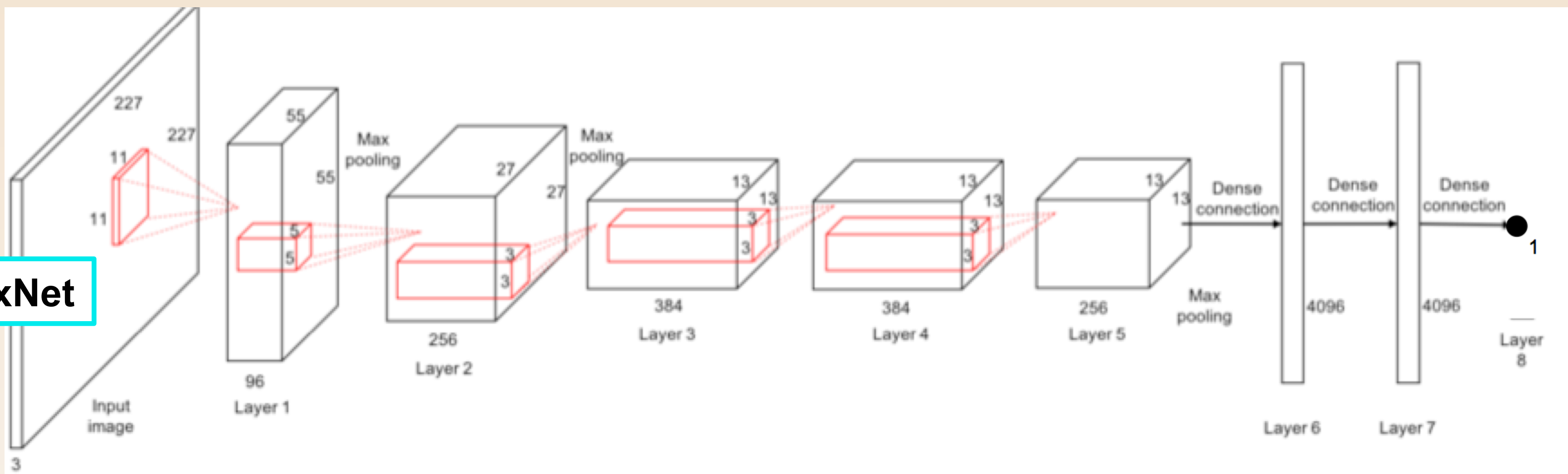


PURDUE UNIVERSITY

Aesthetic Quality Prediction Based on a Convolutional Neural Network

- Our network is very similar to the successful AlexNet.
  - » AlexNet is an 8-layer CNN trained with 1.2 million high-resolution images belonging to 1000 different classes and tested with 150,000 testing images.
    - AlexNet reduced the recognition error rate by 40% compared with the previous best result.
  - » In order to make real-number regression, we replace the last layer's 1000-class softmax classifier with a 1-node neuron.
  - » We initialize our net's parameters with the AlexNet parameters.
    - Its parameters have been well trained to extract image structured features.

AlexNet



PURDUE UNIVERSITY

Data Augmentation

- Make more training data from our 734 training photos to combat overfitting.
- Steps:
  - » Rescale all photos to 256 × 256.
  - » Take 5 patches from each photo. These patches have dimension 227 × 227 and are located at the 4 corners and the center of the image.
  - » Each patch is flipped about the vertical axis.
- Each photo produces 10 training patches.

Positions of upper-left, center, lower-right patches

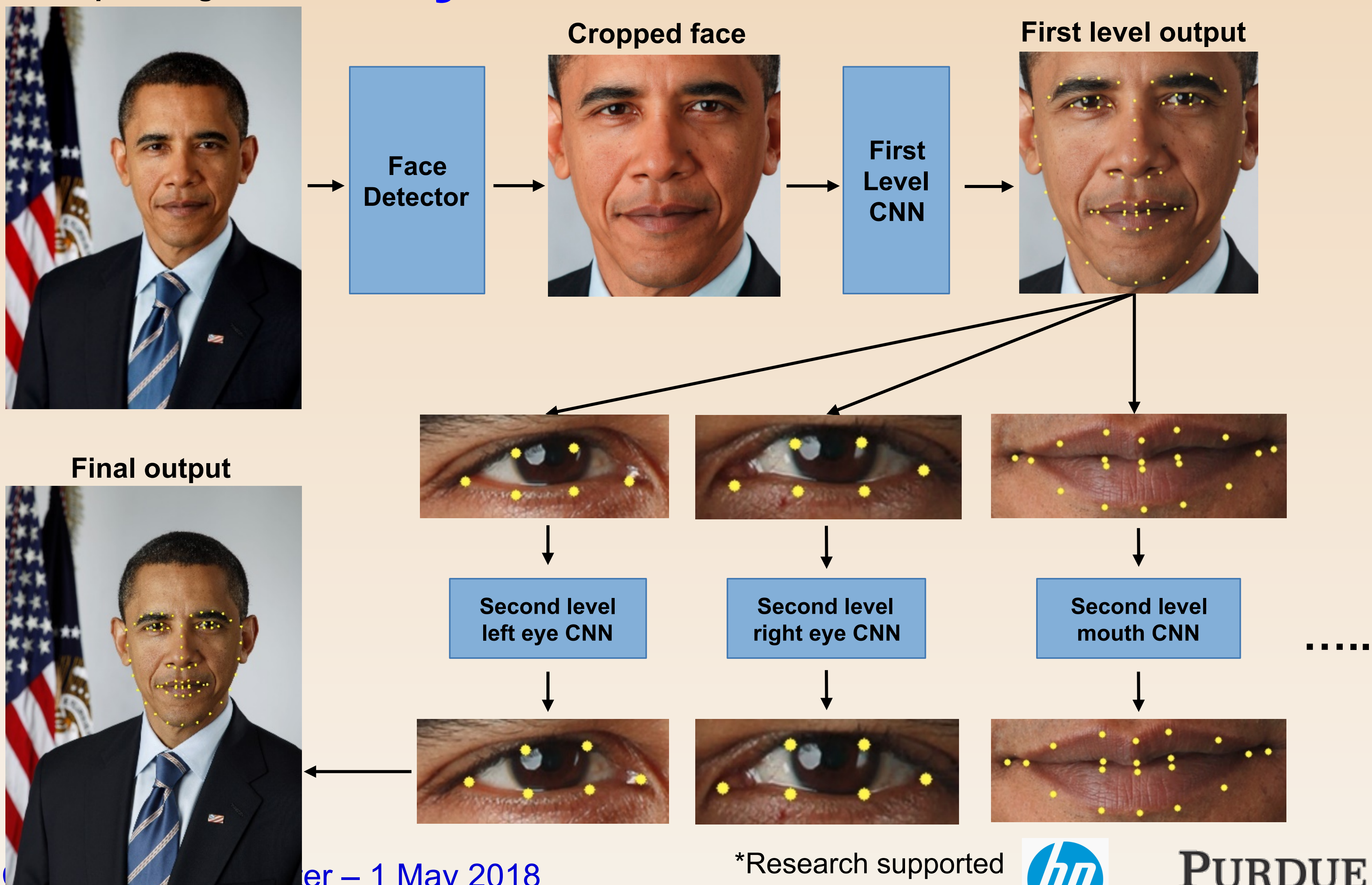
Prediction Accuracy

- We conduct a 5-fold cross-validation to test the prediction accuracy.
- In each fold, we train the net for 100000 iterations.
  - » In each iteration, a batch of the training data is fed into the net and the net parameters are updated by stochastic gradient descent.
  - » Every 200 iterations, we test the accuracy on the testing data set.
  - » The accuracy (or loss since it is the objective of optimization) is calculated as the root mean square error (RMSE) between the ground truth score and the predicted score.
- To verify the importance of initializing our net with AlexNet, we also train a net initialized with random numbers and record the accuracy.

Comparison with SVM Predictor

- RMSE: SVM 1.542, Deep Neural Network 1.530.
  - » Two predictors yield very similar prediction accuracy.
- Which one to use?
  - » It depends.
  - » The deep neural network predictor saves the labor of designing, analyzing, and selecting image features, which is suitable for fast development.
  - » However, the deep neural network model is expensive in storage and computation.
    - Trained deep neural work model needs more than 200 megabytes (MB) storage.
    - In some applications on the mobile platform, the storage and computation could be a bottleneck.

Facial Landmark Detection Using a CNN System Overview*



*Research supported by HP, Inc.

PURDUE UNIVERSITY

Experimental Results

Example landmark predictions using proposed method

Performance Evaluation

Method	68 Point RMSE
SDM (2013)	5.57
CFAN (2014)	5.50
LBF (2014)	4.95
CFSS (2015)	4.73
TCDCN (2014)	4.80
Fan et al. (2016)	4.76
Honari et al. (2016)	4.67
Lai et al. (2016)	4.07
Chen et al. (2017)	3.73
Ours	**3.53**

Comparison of state-of-the-art real time approaches on 300W common test dataset

R. Mao, Q. Lin, and J. Allebach, "CNN Based Facial Landmark Detection," *Imaging and Multimedia Analytics in a Web and Mobile World 2018,* (Part of IS&T Electronic Imaging 2018), J. Allebach, Z. Fan, and Q. Lin, Eds., San Francisco, CA, 28 January -2 February 2018.

Synopsis

- K Nearest Neighbor classification applied to printer forensics
- Extension of K Means to Scalar Sequential Quantization
- Optimal tree-structured piece-wise linear filter for image scaling
- Training-based methods for digital haftoning
- Black-box model for print prediction based on training and linear regression
- Print macrouniformity prediction (Method 1)
- Print macrouniformity prediction (Method 2)
- Fashion photograph aesthetic quality predictor based on SVM and CNN
- Facial landmark detection using CNN
- Logo identification using CNN
- Text field category classification via natural language processing

Scalable Logo Detection and Recognition with Minimal Labeling

Daniel Mas Montserrat[1], Qian Lin[2], Jan Allebach[1], and Edward J. Delp[1]

[1]Purdue University, West Lafayette, IN

[2]HP Inc, Palo Alto, CA

D. Mas, Q. Lin, J. Allebach, and E. Delp, "Scalable Logo Detection and Recognition with Minimal Labeling," *Proceedings of the IEEE 1st International Conference on Multimedia Information Processing and Retrieval*, Miami, FL, 10-12 April 2018.

*Research supported by HP, Inc.

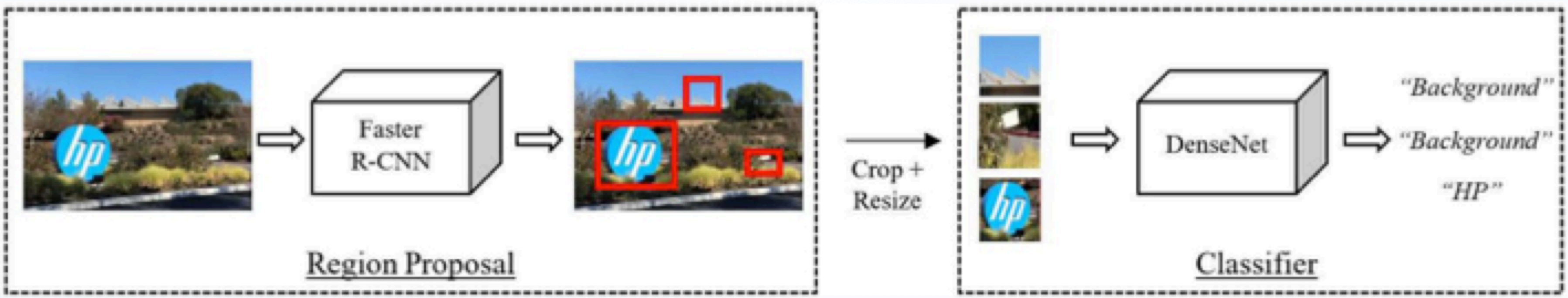
SynthLogo

- **We generate a dataset composed by 280,000 synthetic images including 604 different logos**
- **The 604 logos include well-known brands and products from food, tech, clothing and transportation companies**
- **Images are automatically generated**

Two-Stage Approach: R-CNN

- Faster R-CNN is used as a region proposal

- DenseNet is used as a classifier



Experimental Results

Method	Training Dataset	Number of Classes (Training)	
		32 Logos	604 Logos
Faster R-CNN	SynthLogo	-	47.66%
Faster R-CNN + DenseNet	SynthLogo	74.85%	73.96%
Faster R-CNN + DenseNet	Bootstrap	77.54%	75.12%
Faster R-CNN + DenseNet	SynthLogo + Bootstrap	80.12%	76.89%

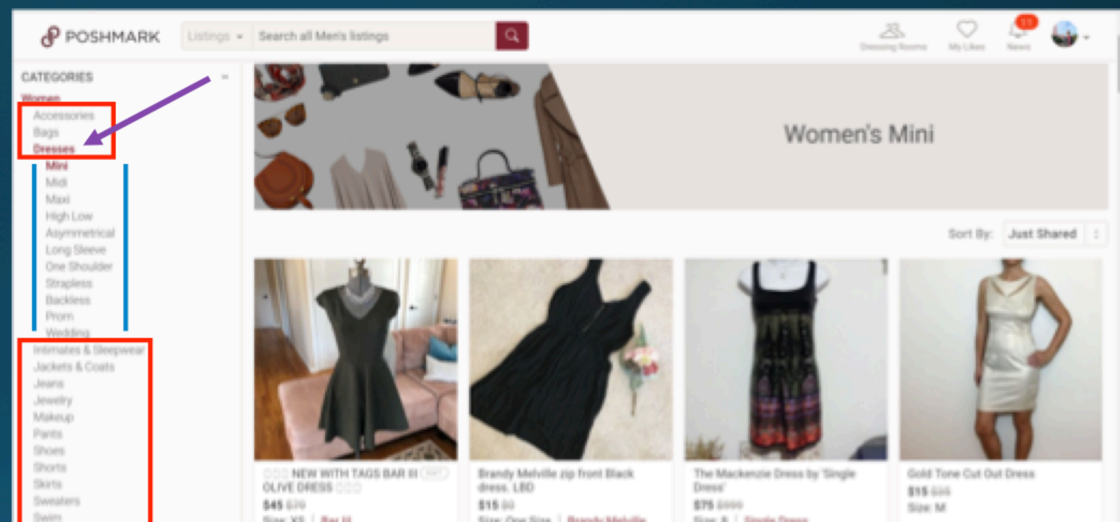
Table1. mAP when using different training data

- **FlickrLogos-32 is used as testing dataset. It contains 32 different logos**
- **Bootstrap images provide a higher mean average precision**
- **The mAP decreases when detecting a larger number of logos**

Application of Natural Language Processing to an Online Fashion Marketplace

Online Fashion Marketplace

- Places to sell and purchase fashion products
- Users are both sellers and buyers
- Users sell items by taking pictures and filling in information on item
- Site has categories, with subcategories within each category
- 15 categories (16 including 'Other')
- Categories range from having 17 subcategories (Ex: Shoes) to 4 subcategories (Ex: Swim)

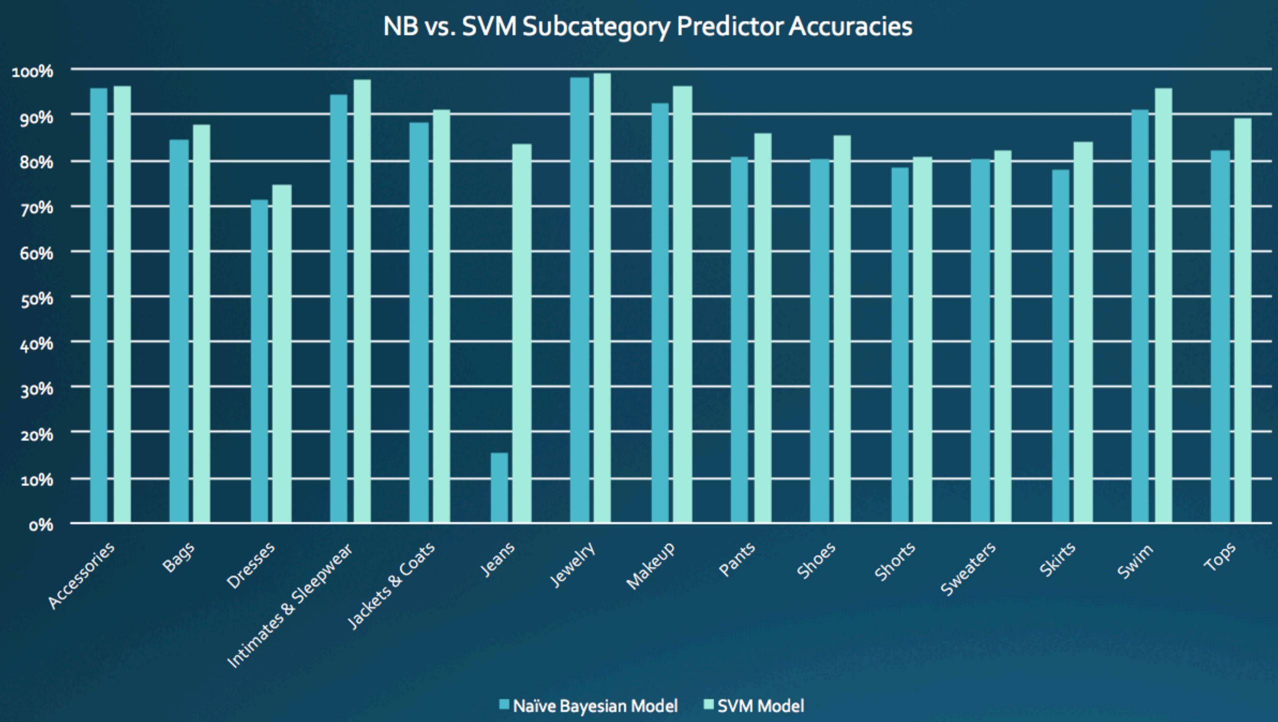


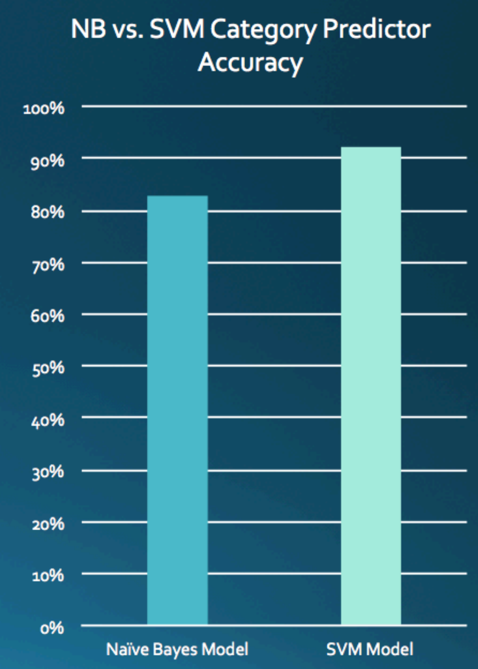
- Surrounds the categories
- On either side of subcategories
- Points to currently selected category

K. Norman, Z. Li, G. Gowala, S. Sundaram, and J. Allebach, "Application of Natural Language Processing to an Online Fashion Marketplace," *Imaging and Multimedia Analytics in a Web and Mobile World 2018,* (Part of IS&T Electronic Imaging 2018), J. Allebach, Z. Fan, and Q. Lin, Eds., San Francisco, CA, 28 January -2 February 2018.

*Research supported by Poshmark, Inc.

Classification Results





Thank you for your interest!