

# Deep Learning approach to Handwriting Comparison

Sargur N. Srihari

University at Buffalo

The State University of New York

[srihari@cedar.buffalo.edu](mailto:srihari@cedar.buffalo.edu)

# Plan of Presentation

1. Variability in Handwriting
2. The Handwriting Comparison Task
3. AI approaches to Handwriting Comparison:
  1. Knowledge-based Approach
  2. Classic Machine Learning Approach
  3. Deep Learning Approach
4. Four Deep Learning Architectures

# Variability in handwriting

referred  
referred  
referred

referred  
referred  
referred

referred  
referred  
referred

refused  
refused  
refused

referred  
referred  
referred

referred  
referred  
referred

referred  
referred  
referred

referred  
referred  
referred

# The Handwriting Comparison Task



# Opinion as a Likelihood Ratio

- $h^0$ : Evidence (with characteristics)  $e$  was generated by *known* (individual with characteristics)  $k$
- $h^1$ : Evidence  $e$  was not generated by  $k$

$$\text{Likelihood Ratio: } LR(k, e) = \frac{p(k, e | h^0)}{p(k, e | h^1)}$$

Probability of *evidence* and *known* (suspect) under Prosecution hypothesis



Probability of *evidence* and *known* (suspect) under Defense hypothesis

# Opinion as a Probability

- Bayesian formulation

- Define prior probabilities  $p(h^0)$  and  $p(h^1)$

Define Prior Odds:  $O_{prior} = \frac{p(h^0)}{p(h^1)}$  Then  $p(h^0) = \frac{O_{prior}}{1 + O_{prior}}$

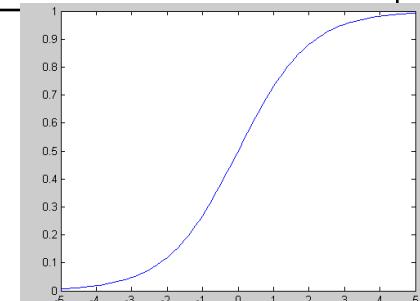
Define  $LR(k, e) = \frac{p(k, e | h^0)}{p(k, e | h^1)}$

Using Bayes rule, Posterior Odds:  $O_{posterior} = O_{prior} \times LR(k, e)$  Then  $p(h^0 | k, e) = \frac{O_{posterior}}{1 + O_{posterior}}$

If  $P(h^0) = P(h^1) = \frac{1}{2}$ ,  $O_{prior} = 1$ ,  $O_{posterior} = LR$

$$p(h^0 | k, e) = \text{sigmoid}(LLR(k, e))$$

- Can fuse results from different features
    - Additive LLRs under independence



# Approaches to Arriving at Opinion

## 1. Knowledge (rule)-based approach

1. Features/Rules are from FDE

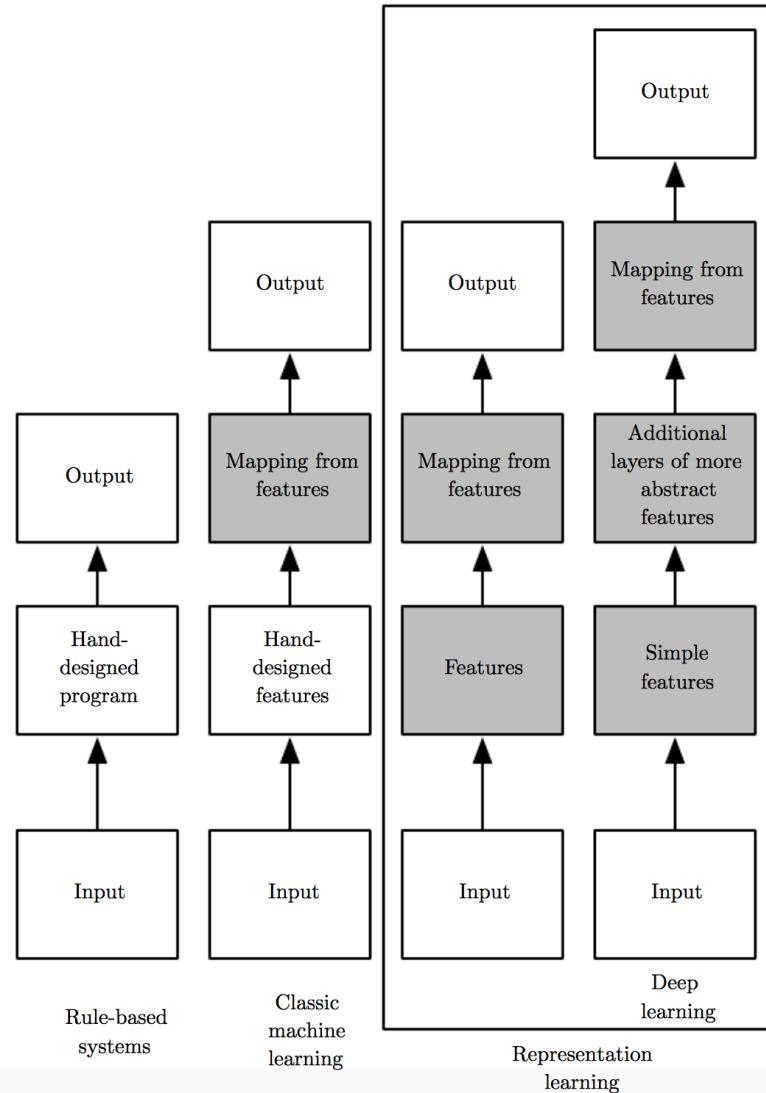
## 2. Classic Machine Learning

1. Features human engineered
2. Decision is learned

## 3. Representation Learning

1. Features/decision learned

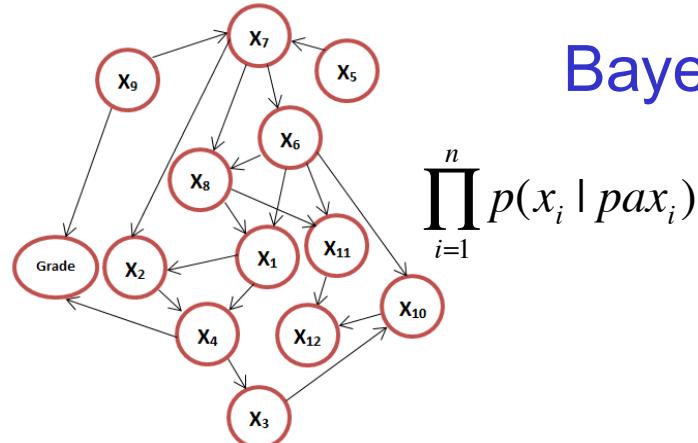
- Shaded boxes involve ML



# Approach 1: Knowledge-based

## Features provided by FDEs Cursive

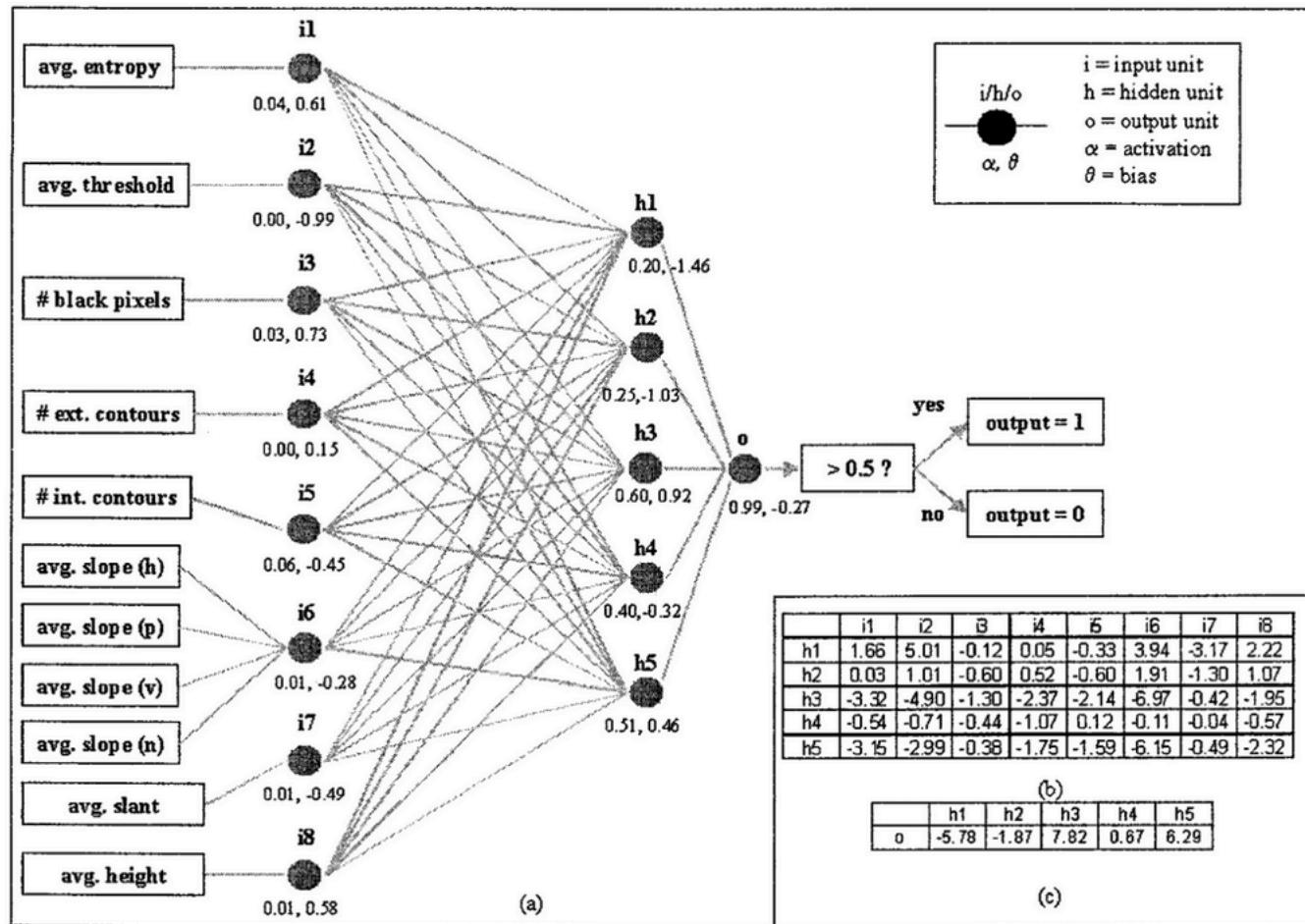
and and and						
Initial stroke of "a"	staff right	a	staff left	a	staff center	a
Formation of "a" staff	tented	a	retraced	a	looped	a
Number of "n" arches	one	n	two	n		
Shape of "n" arches	pointed	n	rounded	n	retraced n combination n	
Location of "n" mid	above base	n	below base	n	at base	n
Formation of "d" staff	tented	d	retraced	d	looped	d
Formation of "d" initial	overhand	d	underhand	d	straight across	d
Formation of "d" terminal	curved up	d	straight	d	curved down d no obvious end stroke	d
Symbol	unusual	d		symbol i +		
a-n relationship	a taller	and	a equal	and	a smaller and	
a-d relationship	a taller	and	a equal	and	a smaller and	
n-d relationship	n taller	and	n equal	and	n smaller and	



## Bayesian Network of feature differences

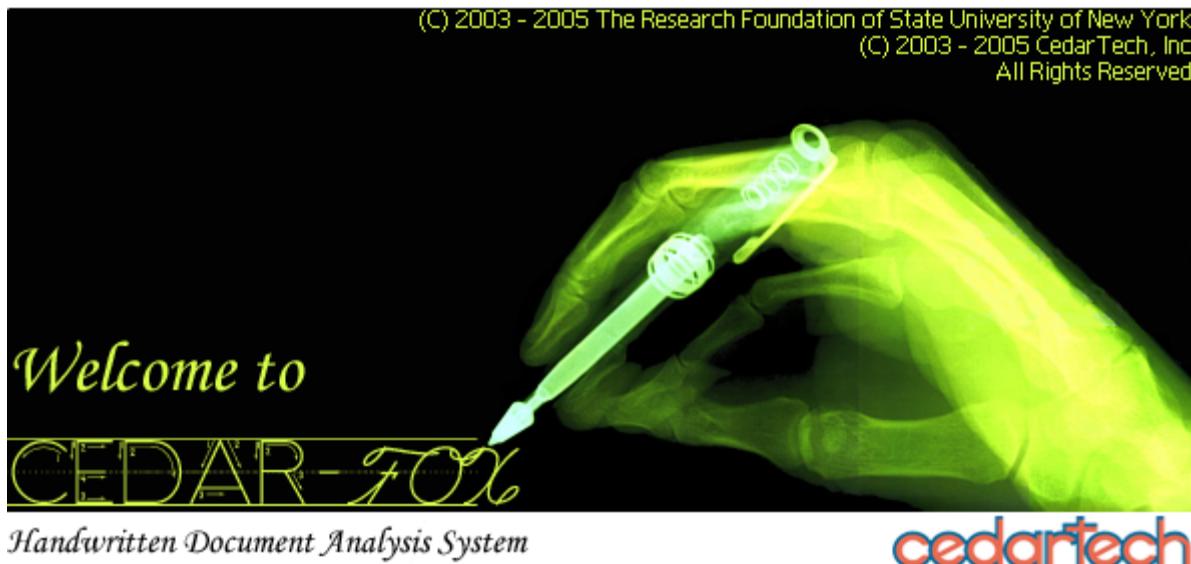
Results: Accuracy  
Cursive : 75%

# Approach 2: Simple machine learning



# Approach 2: CEDARFOX software

- Writer Verification/Identification
  - Computes LLR based on human engineered features
  - Similarities based on Document Properties
    - Line Structure, Writer Characteristics



# CEDAR-FOX: Side-by-side Comparison of Extended Writing

Known

From  
Jim Elder  
829 Loop Street, Apt. 300  
Allentown, New York 14707

To  
Dr. Bob Grant  
602 Queenberry Parkway  
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack. It all started around six months ago while attending the "Rubeg" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank You!

Jim



Questioned

November 10, 1994

From  
Jim Elder  
829 Loop Street, Apt. 300  
Allentown, New York 14707

To  
Dr. Bob Grant  
602 Queenberry Parkway  
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeg" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several question, x-rays and blood tests later, were told it was just exhaustion.

# Result of Comparsion

**CEDAR-FOX - 0002bl.png -- Known Document**

File View Features Search Tools Recognition Window Help

0002bl.png - Known Document

However, the extra hours affected her health, walking through the show she passed out. We rushed her to the hospital, and several questions, X-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!

Jim

**0001bu.png -- Questioned Document**

From: Jim Elder Nov 10, 1999

To: Dr. Bob Grant

829 Loop Street, Apt 300  
Altenau, New York 14700

602 Greenberry Parkway  
Omar, West Virginia 0

We were referred to your Center. This is regarding my son. It all started around six years ago. Organizing each of the Alumni Association,

**Calibrated Score**

Opinion

Given that the documents compared in either questioned or known had

Line  Paragraph  Half Page  Full Page

of text and that the documents compared had

Same Content  Different Content

the opinion is:

Identified As Same  
 Highly Probable Same  
 Probably Did  
 Indication Did  
 No Conclusion  
 Indication Did Not  
 Probably Did Not  
 Highly Probable Did Not  
 Identified As Different

OK

**Document Features**

File

Questioned: 0001bu.png  
Known: 0002bl.png

Macro Features

Log Same: -7.71  
Log Diff: -1.74  
LLR: -5.96

View Details

Micro Features of Characters

Characters to be used

Automatically Recognized  Word-Truthed  Manually Cropped  None

Automatically Recognized

Use these features?

Chars from Question: 68  
Chars from Known: 105  
Number of Distances: 476

Log Same: 27.87  
Log Diff: 63.95  
LLR: -36.07

Word-Truthed

Use these features?

Chars from Question: 162  
Chars from Known: 142  
Number of Distances: 1328

Score Same: 19.28  
Score Diff: 27.24  
Score: -7.96

Manually Cropped

Use these features?

Number of Distances: 0  
Log Same: 0  
Log Diff: 0  
LLR: 0

View Details

Style Features

Bigram

Use these features?

Bigram LLR: -12.19 View Details  
"th" LLR: 3.51 View Details

Word

Use Word Features?

LLR value: -15.13  
No. of Words: 16 View Details

View Details

Document Similarity

Type of Document

Line  Paragraph  Half Page  Full Page

Type of Content

Same Content  Different Content

Log Likelihood Ratio: -41.25

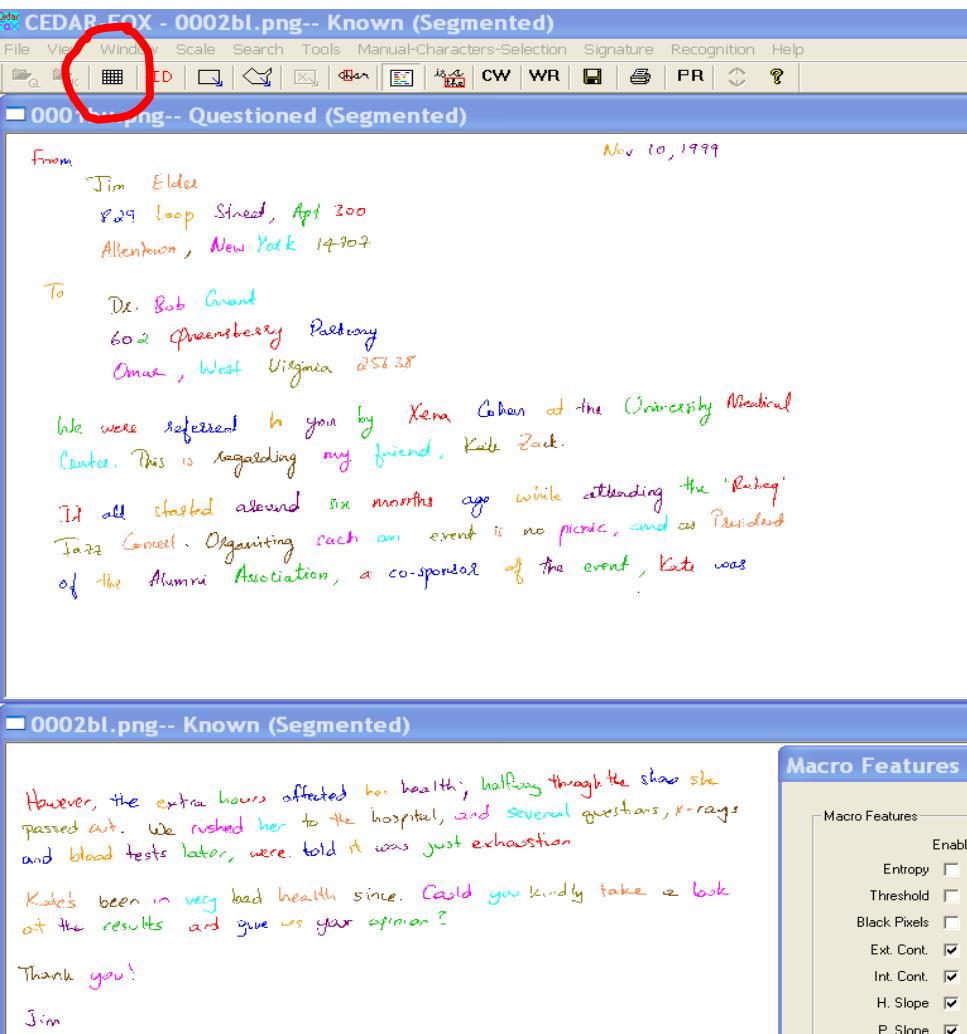
Same Writer? No Interpretation

0 % of the same writer test cases had a LLR value less than the LLR value of this pair.  
41.99 % of the different writer test cases had a LLR value less than the LLR value of this pair.

**Feature Comparison Table**

**Strength of Evidence**

# Human Engineered Characteristics



**Document Features**

**Pictorial Attribute Scores**

**Letter Formation Scores**

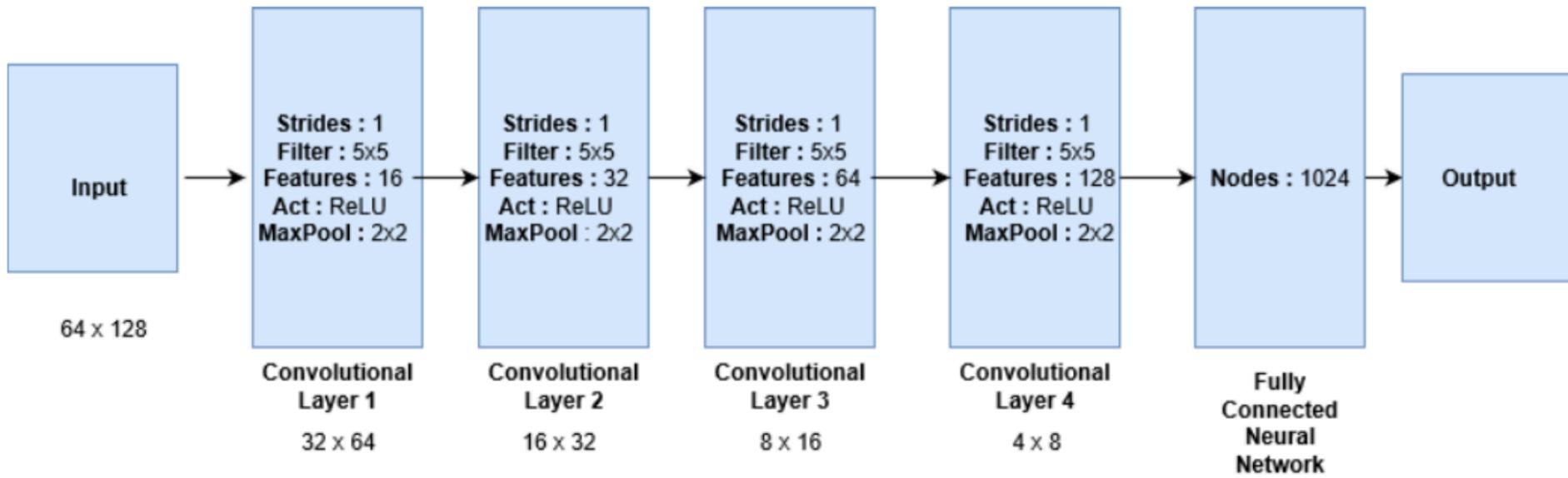
This screenshot shows the 'Document Features' window of the software. It includes sections for 'Micro Features of Characters' (with radio buttons for 'Automatically Recognized', 'Word-Truthed', 'Manually Cropped', and 'None'), 'Macro Features' (with numerical values like Log Same: -8.32, Log Diff: -1.75, LLR: -6.57), and 'Document Similarity' (with a Log Likelihood Ratio of -6.57). A yellow box labeled 'Pictorial Attribute Scores' highlights the 'Macro Features' section, and another yellow box labeled 'Letter Formation Scores' highlights the 'Document Similarity' section. Red arrows point from these labels to their respective sections in the window.

**Macro Features**

**Automatically Identified**

	Weight	Q-Rec	K-Rec	
0	0.1140	10	16	0.68
1	0.0376	5	11	0.67
2	0.5645	3	1	0.57
3	0.3617	2	0	0.00
4	0.6067	1	0	0.00
5	0.5435	1	2	0.67
6	0.5389	3	1	0.58
7	0.3342	1	1	0.59
8	0.3662	3	0	0.00
9	0.4833	4	0	0.00
a	0.5239	1	1	0.60
b	0.9462	3	2	0.66
c	0.2574	3	1	0.70
x	0.4114			

# Approach 3: Deep Learning Architecture 1



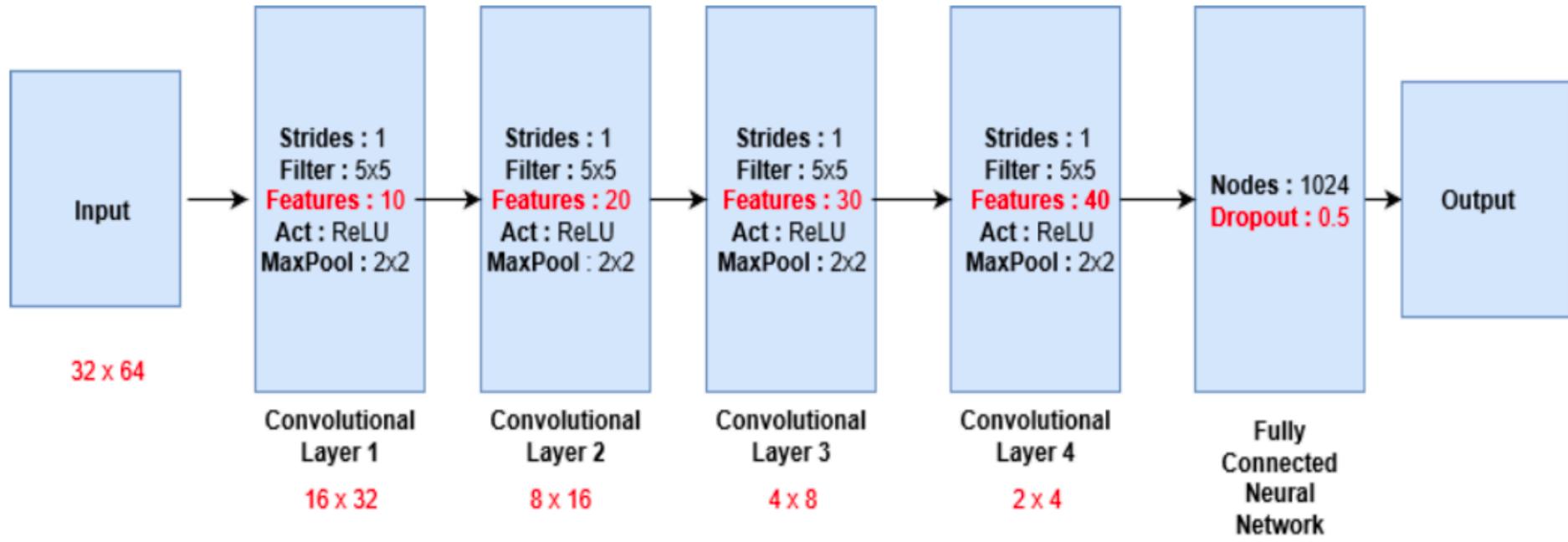
Stack images and feed to a 4 layer CNN (feature extractor) and a FCN for classification

## Results

Type -1 : 59%

Type 2 : 65%

# Deep Learning Architecture 2

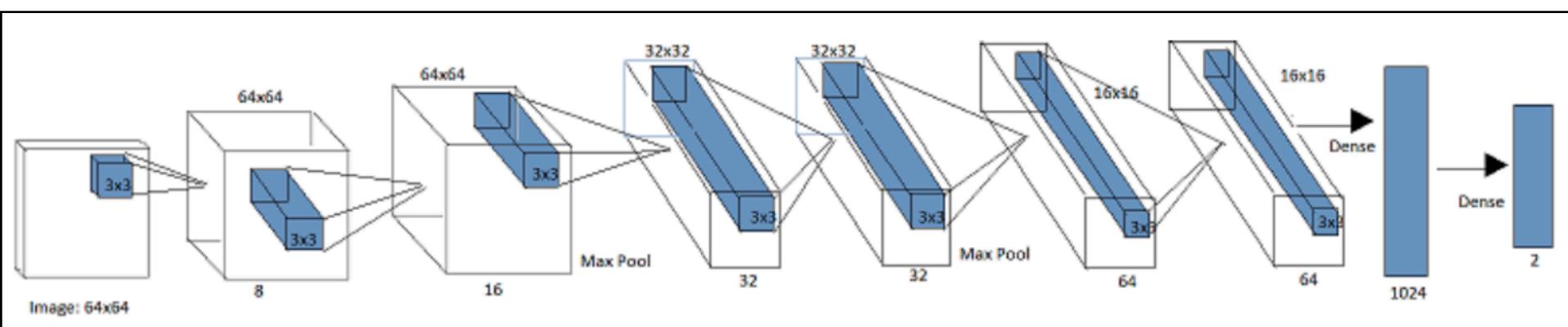


First resize images and standardize on image padding and strides  
Include regularization: dropout in the fully connected layer

## Results

Type -1 : 85%  
Type -2 : 78%

# Deep Learning Architecture 3



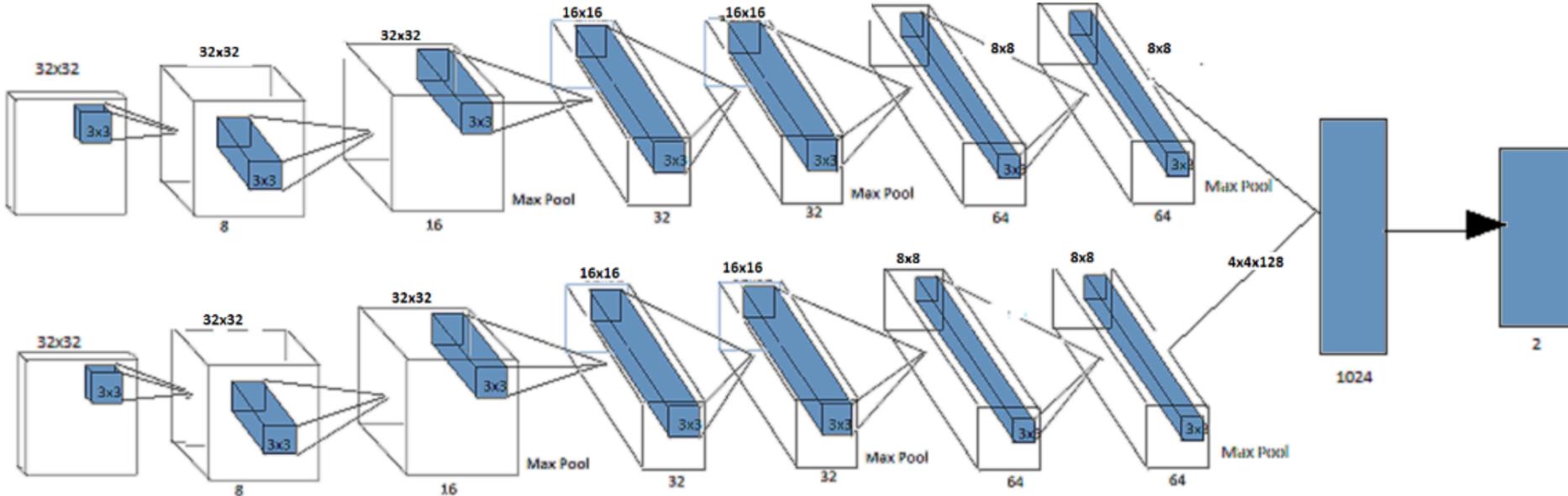
Introduce maxpooling and batch normalization

Results:

Type -1 : 95%

Type -2 : 84%

# Deep Learning Architecture 4



Using same architecture as above but split into two channels

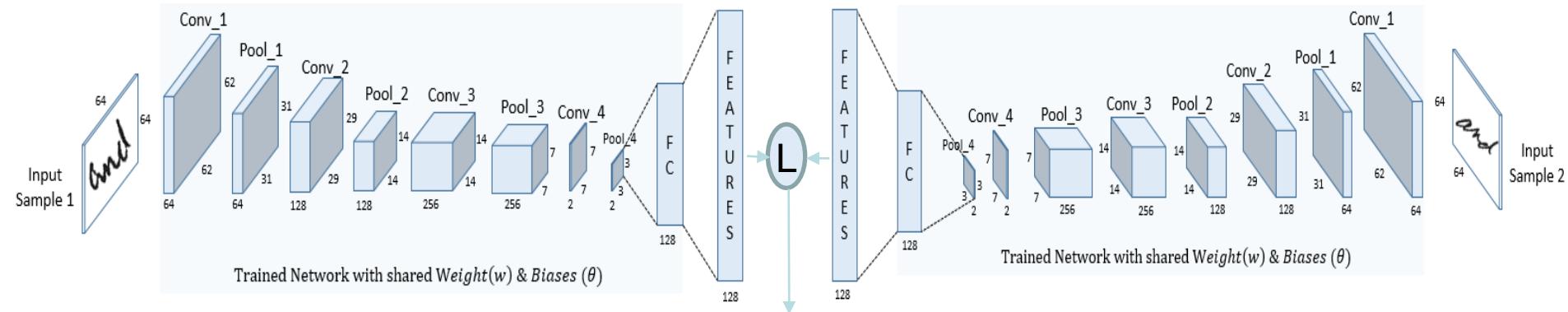
Results:

Type -1 : 96%

Type -2 : 89%

# Architecture 5: One Shot Twin Network

<https://github.com/mshaikh2/DeepLearningHandwritingForensics>



## Algorithm 1: One Shot Twin Network

**Data:** Known & Query Image

**Result:** Writer Verification

Known Image:  $X_i$  & Query Image:  $X_j$  ;

Dissimilarity Model:  $M_{w\_diss}$ ; Weights of  $M_{w\_diss}$ :  $W\_dis$

Similarity Model:  $M_{w\_sim}$ ; Weights of  $M_{w\_sim}$ :  $W\_sim_{ij}$  ;

Contrastive Loss:  $C_{Loss}$ ;

for  $X_i \dots X_j$  do

    Update  $W\_diss_{ij}$  for  $M_{w\_diss}$  assuming samples classifying in different classes [0,1] ;

    Update  $W\_sim_{ij}$  for  $M_{w\_sim}$  assuming samples classifying in the same classes [1] ;

end

for Model in [ $M_{w\_diss}$  ,  $M_{w\_sim}$ ] do

    for Datapoint in [ $X_i \dots X_j$ ] do

        | Extract features  $F_{Datapoint}[0..127]$  for Datapoint from Model;

    end

    Calculate:  $C_{Loss} (F_i[0..127], F_j[0..127])$

end

### Log likelihood Contrastive Similarity

$$\ln[C_{Loss}(M_{w-sim})/C_{Loss}(M_{w-dis})]$$

$$C_{Loss}(M_{w-dis}) = 0.5 * \{ \max(0, m - D_w) \}^2$$

$$C_{Loss}(M_{w-sim}) = 0.5 * \{ D_w \}^2$$

# Comparison of Two Approaches

TYPE I: Intra Class Accuracy

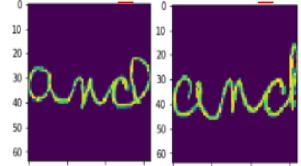
(LLR)

Known Quest

CEDA  
R-FOX

OSTN

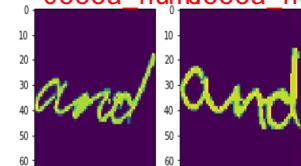
0034c\_num0034c\_num2



-19.  
X

0.308  
✓

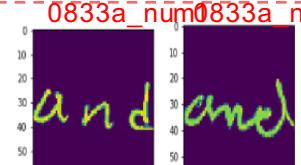
0883a\_num0883a\_num3



-  
13.2  
4

-  
0.59  
X

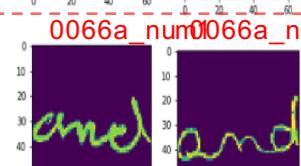
0833a\_num0833a\_num2



-  
22  
45

0.0  
54  
✓

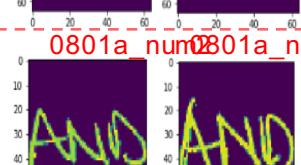
0066a\_num0066a\_num2



-  
5.8  
4

0.76  
✓

0801a\_num0801a\_num2



8.9  
2  
✓

0.1  
9  
✓

TYPE II: Inter Class Accuracy

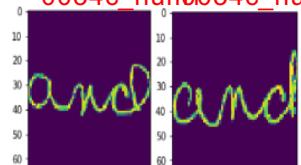
(LLR)

Known Quest

CEDA  
R-FOX

OSTN

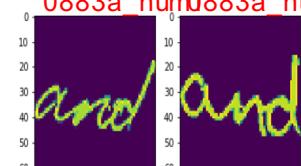
0066a\_num0801a\_num1



-  
33.  
✓

-0.62  
✓

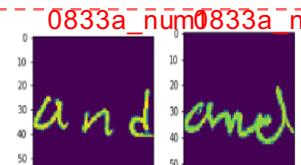
0066a\_num0833a\_num2



-16.  
✓

-0.707  
✓

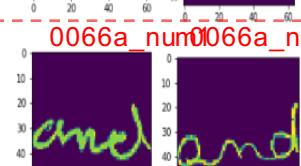
0034c\_num0833a\_num1



-27.  
✓

-1.1  
34  
✓

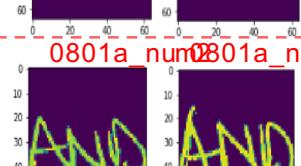
0801a\_num0833a\_num1



-  
25.9  
✓

1.0  
8  
✓

0034c\_num0833a\_num1

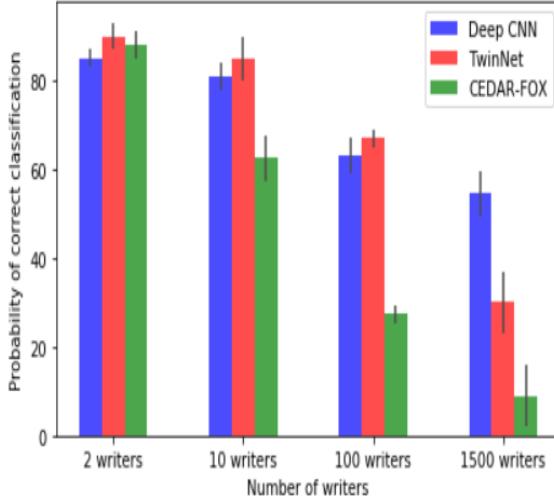


-  
30  
69  
✓

-1.06  
✓

Performance with no of writers

Comparison of various techniques



Comparison	Type I	Type 2
------------	--------	--------

CEDAR-FOX	77%	88%
-----------	-----	-----

OSTN	90%	96%
------	-----	-----

Negative values measure dissimilarity, Positive values measure similarity.

# Summary

- Compared performance of three AI approaches for the task of handwriting comparison
  1. Knowledge-based: Features specified by FDEs
  2. Simple machine learning: Features engineered
  3. Deep learning: Features learnt automatically
- Observed progressive performance improvement over the three approaches