Class scores

TRAINABLE CLASSIFIER MODULE

Feature vector

FEATURE EXTRACTION MODULE

Raw input









INPUT

C1: feature maps

C3: f. maps 16@10x10

S4: f. maps 16@5x5



32x32

6@28x28

S2: f. maps 6@14x14

C5: layer 120

F6: layer 84

OUTPUT 10

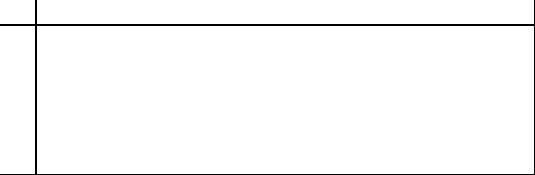
Full connection Gaussian connections

Convolutions

Subsampling

Convolutions Subsampling

Full connection



! " # $ % & ’ ( ) \* + , − . /



0 1 2 3 4 5 6 7 8 9 : ; < = > ?



@ A B C D E F G H I J K L M N O



P Q R S T U V W X Y Z [ \ ] ^ \_

‘ a b c d e f g h i j k l m n o



p q r s t u v w x y z { | } ~























Error Rate (%)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

5%

4%

3%

2%

1% Test

0% Training  

0 4 8 12 16 20      

Training set Iterations

1.8

1.6

1.4

1.2

1

Error Rate (%)

distortions)

Test error (no

4−>6 3−>5 8−>2 2−>1 5−>3 4−>8 2−>8 3−>5 6−>5 7−>3

9−>4 8−>0 7−>8 5−>3 8−>7 0−>6 3−>7 2−>7 8−>3 9−>4

8−>2 5−>3 4−>8 3−>9 6−>0 9−>8 4−>9 6−>1 9−>4 9−>1

9−>4 2−>0 6−>1 3−>5 3−>2 9−>5 6−>0 6−>0 6−>0 6−>8

0.8

0.6

0.4

0.2

0 0

###### Test error

(with distortions)

Training error (no distortions)

10 20 30 40 50 60 70 80 90 100

###### Training Set Size (x1000)

4−>6 7−>3 9−>4 4−>6 2−>7 9−>7 4−>3 9−>4 9−>4 9−>4

8−>7 4−>2 8−>4 3−>5 8−>4 6−>5 8−>5 3−>8 3−>8 9−>8

1−>5 9−>8 6−>3 0−>2 6−>5 9−>5 0−>7 1−>6 4−>9 2−>1

2−>8 8−>5 4−>9 7−>2 7−>2 6−>5 9−>7 6−>1 5−>6 5−>0

4−>9 2−>8

Linear [deslant] Linear

−−−− 12.0 −−−−>

−−−− 8.4 −−−−>

−−−− 7.6 −−−−>

2.4

3.3

3.6

1.1

1.1

1

0.8

4.7

3.6

1.6

4.5

3.8

3.05

2.5

2.95

2.45

1.7

1.1

1.1

1.1

0.95

0.8

0.7

Pairwise

K−NN Euclidean 5

[deslant] K−NN Euclidean 40 PCA + quadratic 1000 RBF + linear

[16x16] Tangent Distance

SVM poly 4

RS−SVM poly 5 [dist] V−SVM poly 9

28x28−300−10 [dist] 28x28−300−10

[deslant] 20x20−300−10

28x28−1000−10 [dist] 28x28−1000−10 28x28−300−100−10

[dist] 28x28−300−100−10 28x28−500−150−10

[dist] 28x28−500−150−10

[16x16] LeNet−1

LeNet−4 LeNet−4 / Local LeNet−4 / K−NN

LeNet−5 [dist] LeNet−5

[dist] Boosted LeNet−4

0 0.5

1 1.5

2 2.5

3 3.5

4 4.5 5

[deslant] K−NN Euclidean [16x16] Tangent Distance

3.2

1.4

0.5

1.6

1.8

3.7

1.9

1.8

8.1

SVM poly 4

[deslant] 20x20−300−10

[16x16] LeNet−1

LeNet−4 LeNet−4 / Local LeNet−4 / K−NN

[dist] Boosted LeNet−4

0 1 2 3 4 5 6 7 8 9

Linear Pairwise

4

36

123

795

267

469

100

260

−−−− 20,000 −−−−>

−−−− 10,000 −−−−>

401

460

|  |  |  |  |
| --- | --- | --- | --- |
| [deslant] K−NN Euclidean  40 PCA+quadratic | −−−− 24,000 −−−−>  39 |  | |
| 1000 RBF |  |  | 794 |
| [16x16] Tangent Distance  SVM poly 4  RS−SVM poly 5 | −−−− 20,000 −−−−>  −−−− 14,000 −−−−> |  | 650 |
| [dist] V−SVM poly 9 | −−−− 28,000 −−−−> |  |  |

[deslant] 20x20−300−10

28x28−1000−10

28x28−300−100−10

28x28−500−150−10

[16x16] LeNet−1

LeNet−4 LeNet−4 / Local LeNet−4 / K−NN

LeNet−5 Boosted LeNet−4

0 300 600 900

Linear Pairwise

4

35

123

795

267

469

3

17

−−− 24,000 −−−>

−−− 24,000 −−−>

60

51

|  |  |  |  |
| --- | --- | --- | --- |
| [deslant] K−NN Euclidean  40 PCA+quadratic | −−− 24,000 −−−>  40 |  |  |
| 1000 RBF |  |  | 794 |
| [16x16] Tangent Distance  SVM poly 4  RS−SVM poly 5 | −−− 25,000 −−−>  −−−− 14,000 −−−−> |  | 650 |
| [dist] V−SVM poly 5 | −−−− 28,000 −−−−> |  |  |

[deslant] 20x20−300−10

28x28−1000−10

28x28−300−100−10

28x28−500−150−10

[16x16] LeNet 1

LeNet 4 LeNet 4 / Local LeNet 4 / K−NN

LeNet 5

Boosted LeNet 4

0 300 600 900

W1 W2

Input

X3

F3(X3,X4)

Loss Function

F1(X0,X1,W1)

X5

Z

X4

F2(X2,W2)

F0(X0)

X1

E

X2

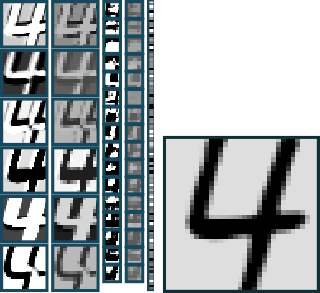
D

Desired Output

**C1 S2 C3 S4 C5**



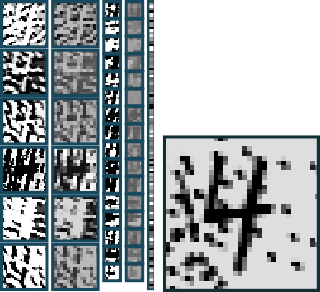
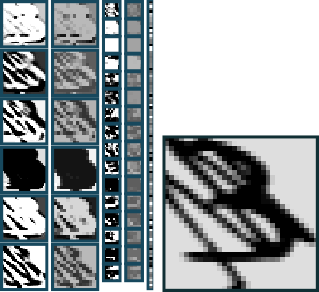
**4 4 4**



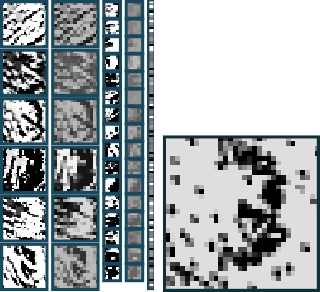
**Output**

**F6**

**4 3 8**



**4 3 3**



**(a)**

LLaayyeerr

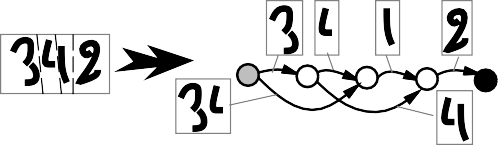
LLaayyeerr

Graph Transformer



**(b)**

Graph Transformer



**Viterbi Penalty**





2

3

4

class label

character recognizer penalty for each class

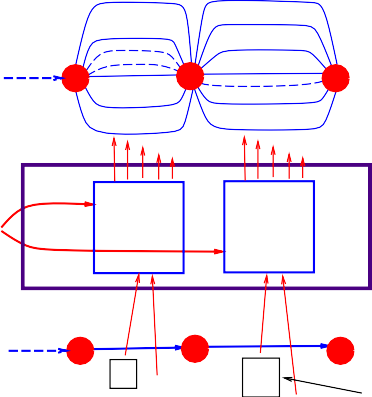
**G**vit

**T** vit

###### Viterbi Path

PIECE OF THE

INTERPRETATION GRAPH



"0" 6.7

"1" 10.3

"8" 0.3

"9" 12.5

"0"

"1"

"2"

"3"

"8"

"9"

7.9

11.2

6.8

0.2

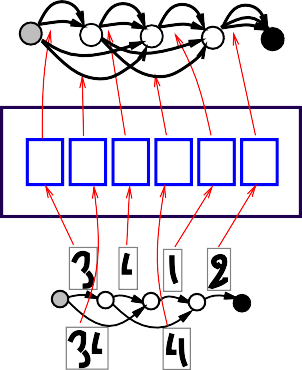
13.5

8.4

**G**int

3 3 1

2 4 4



34 4

1

###### 2 Interpretation

4 Graph

3

Character Recognizer

Viterbi

Transformer

##### 8

W

Character Recognizer

##### 3

candidate

**T** rec

**G**seg

**NN NN NN NN NN NN**

Recognition Transformer

###### Segmentation Graph

0.1

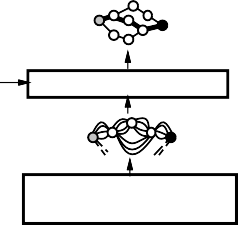
PIECE OF THE 0.5

SEGMENTATION GRAPH

segment image

penalty given by the segmentor

**Constrained Viterbi Penalty**





Ccvit

**Best Constrained Path**

Gcvit

**Constrained Interpretation Graph**

**Viterbi Transformer**

Gc

**Desired Sequence Path Selector**

**Interpretation Graph**

Gint

**Recognition Transformer**

Loss Function



[0.1](+1)

[0.7](+1)

+



−

[0.6](−1)

+

+

3 [0.1](+1)

4 [0.6](+1)

G

3 [0.1](−1) 4 [0.4](−1) 1 [0.1](−1)

**Viterbi Tansformer**



Gcvit

vit

3 [0.1](+1)



Gc

4 [2.4](0)

**Viterbi Transformer**

**"34"**

Desired Answer

3 [3.4](0) 4 [0.6](+1)

**Path Selector**

3 [0.1](0)

5 [2.3](0)

3 [3.4](0)

4 [4.4](0)

4 [0.4](−1)

2 [1.3](0)

1 [0.1](−1)

4 [2.4](0)

4 [0.6](+1)

9 [1.2](0)

Interpretation Graph

**G**int

# W

## Neural Net Weights

#### (−1) (+1) (−1)

**NN**

**NN**

4

**NN**

4

**NN**

1

**NN**

Recognition Transfomer

**T** rec

## Segmentation Graph

**G**seg



**Segmenter**

**Cdforw**

**Edforw**

**+ −**



**Cforw**

**Forward Scorer**

**Constrained Interpretation Graph**

**Desired Sequence**



**Recognition Transformer**

Gc

**Path Selector**

**Forward Scorer**

**Interpretation Graph**

Gint



**"U"**

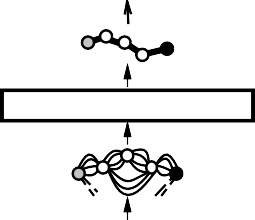


**$**

**Recognizer**

Viterbi Answer

**Viterbi Graph**



**Viterbi Transformer**

**Interpretation Graph**

Character Model Transducer



**Compose**

**SDNN Output**

**S....c.....r......i....p. t**

**s....e.....n.....e.j. o.T**

**5......a...i...u......p. f**

**SDNN**

**Transformer**

**C1 C3 C5**



**2345**

**Compose + Viterbi**

**2 3 3 4 5**



**F6**

**Input**

**Answer SDNN**

**Output**

**540**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1114** | | | | | | | | |
| **1 1 1 4 4 1** | | | | | | | | |
|  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **5 5 4 0** | | | | | | | | |
|  |  |  |  |  |  |  |  |  |

**Answer**



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **678** | | | | | | | | |
| **6 7 7 7 8 8** | | | | | | | | |
|  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **3514** | | | | | | | | |
| **3 5 5 1 1 4** | | | | | | | | |
|  |  |  |  |  |  |  |  |  |

**SDNN**

**output**



**F6**

**Input**

**Cdforw**

**Edforw**

### + −



**Cforw**

**Forward Scorer**

**Forward Scorer**

**Constrained Interpretation Graph**



Gc

**Desired Sequence**

**Path Selector**

**Interpretation Graph**

Character Model Transducer



**Compose**

Gint

**SDNN Output**

**S....c.....r......i....p. t**

**s....e.....n.....e.j. o.T**

**5......a...i...u......p. f**

**SDNN**

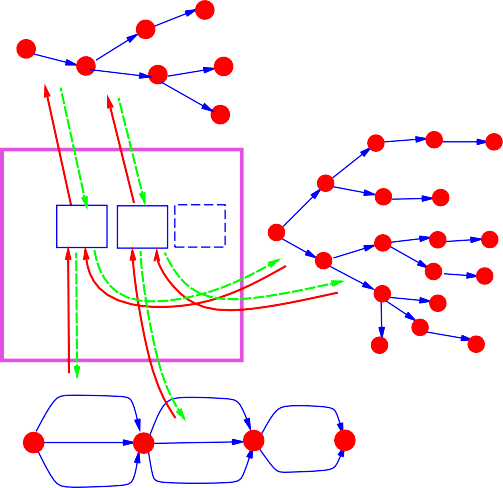
**Transformer**

**interpretation graph**

**"t" 0.8**

**interpretations: cut (2.0)**

**cap (0.8)**

**cat (1.4)**

**"c" 0.4**

**"u"0.8**

**"a" 0.2**

**"p" 0.2**

**"t"**

**0.8**

**grammar graph**

**"r" "n"**

**match & add**

**Graph Composition**

**match & add**

**match & add**

**"b"**

**"c"**

**"a"**

**"u"**

**"u"**

**"a"**

**"t"**

**"t"**

**"r"**

**"p"**

**"e"**

**"e"**

**"t" "r"**

**"d"**

**"c" 0.4**

**"x"**

**0.1**

**"p" 0.2**

**"o" 1.0**

**"a" 0.2**

**Recognition Graph**

**"d" 1.8**

**"u" 0.8**

**"t" 0.8**

**Viterbi Graph**

**Beam Search Transformer**

**Interpretation Graph**

Language Model

**Compose**

**Recognition Graph**

**Recognition Transformer**

**AMAP Graph**

"Script"





**Viterbi Graph**

**Interpretation Graph**

**Beam Search Transformer**

Language Model

**Compose**

**Recognition Graph**

Character Model

**Compose**

**SDNN Output**

"Script"

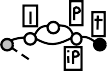


**AMAP Computation**

**SDNN**

**Transformer**

**Segmentation Graph**



**AMAP**



**Segmentation Transformer**

**AMAP Computation**

**Normalized Word**



**Word Normalization**

**Normalized Word**



**Word Normalization**



**SDNN/HMM**



12.4

8.2

8.5

6.3

25K Word Lexicon 2

1.4

No Language Model

No Language Model

no global training with global training

**HOS**

no global training with global training

HOS

no global training with global training

0 5 10 15

Viterbi Answer

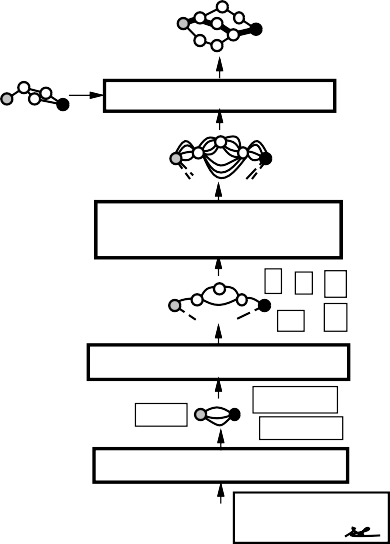
**Best Amount Graph**

**Viterbi Transformer**

**Interpretation Graph**

**"$" 0.2**

**"\*" 0.4**



**"3" 0.1**

**.......**

**Grammar Recognition Graph**

**Segmentation Graph**

**Compose**

**"$" 0.2**

**"\*" 0.4**

**"3" 0.1**

**"B" 23.6**

**.......**

**Recognition Transformer**

**$ \* 3**

**\*\* 45**

**Segmentation Transf.**

**Field Graph**

**45/xx**

**$ \*\*\* 3.45**

**$10,000.00**

**Field Location Transf.**

**Check Graph**

**2nd Nat. Bank**

**not to exceed $10,000.00 three dollars and 45/xx**

**$ \*\*\* 3.45**

**Edforw**



**Cdforw**

**+**

**−**

**Cforw**

**Forward Scorer**

**Forward Scorer**

**Viterbi Answer**



**Path Selector**

**Interpretation Graph**























