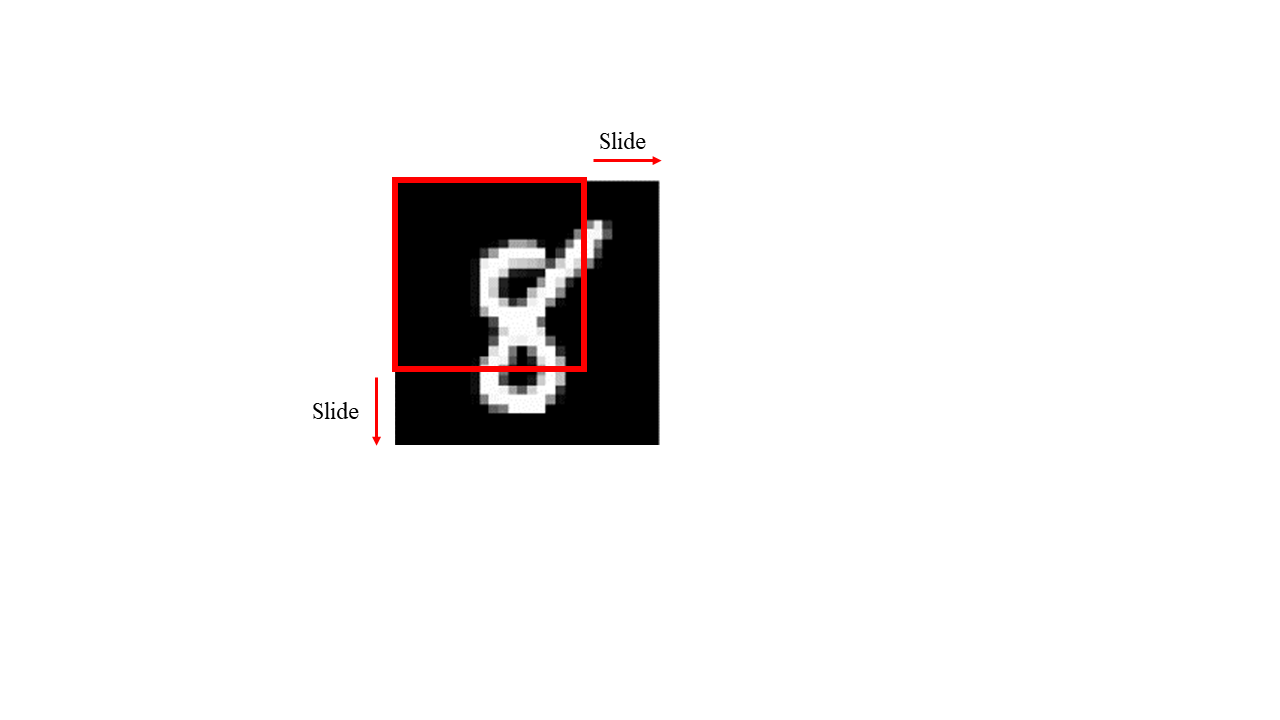
**Group Learning for improving performance of classifiers**

**Preparation of the group data using fixed-size window**

For group learning, an image is partitioned as a group of sub-images using a window with a fixed size. For instance (see Fig. 1), a digit eight image is partitioned using a window of 20\*20 pixels. The window is initially placed at the upper-left corner of the image to extract the sub-image within it (for 20\*20 pixels). After each extraction, the window slides to one pixel to the right/down direction, and then extracts the next sub-image. The extraction keeps until there are no pixels enough to move. Hence, 64 sub-images of the digit eight are extracted for group learning. For the cropped face data with more pixels (160\*160), the sub-images are extracted through the same process, except the slide is for 30 pixels.

****

**Fig. 1.** Partition the image for group learning

**SVM**

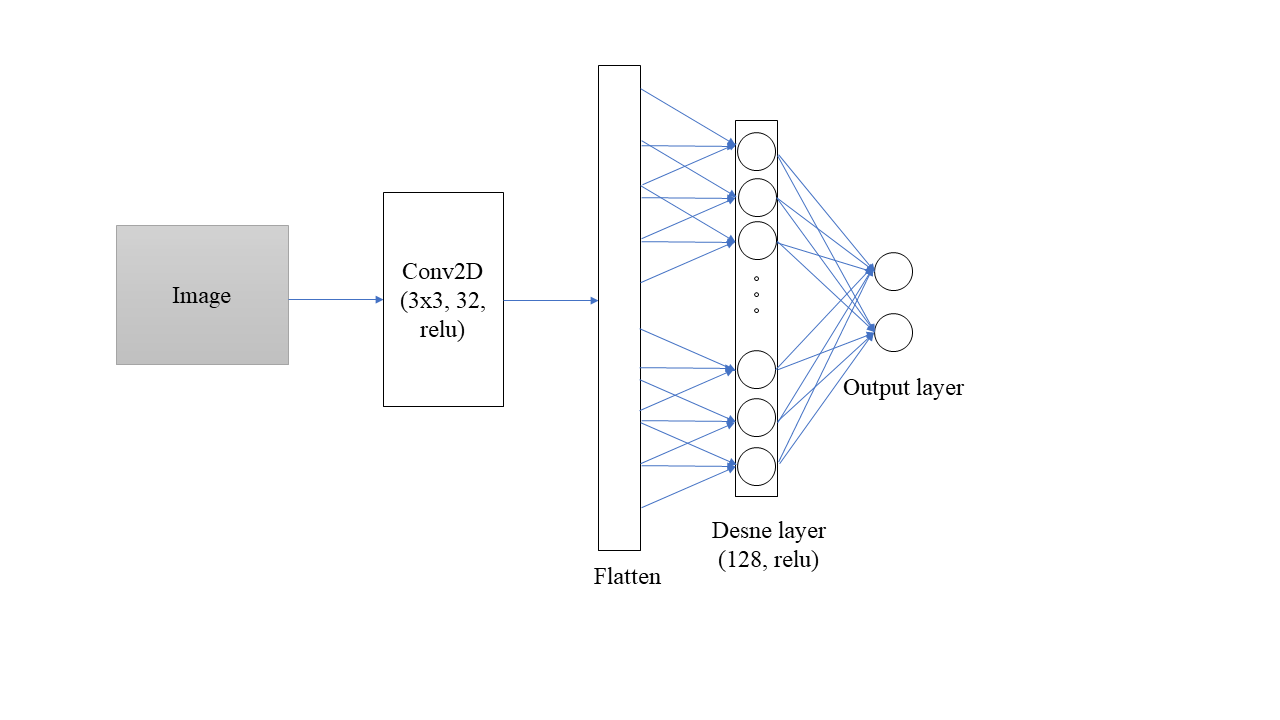
Linear SVM is used in this study for high-dimensional sparse data. LIBLINEAR package (Fan, Chang, Hsieh, Wang, & Lin, 2008) is used for estimating SVM classifier.

**CNN model**

The CNN model used in this study was built for recognizing handwritten digits (<https://www.tensorflow.org/tutorials/quickstart/advanced>). The network is revised for binary classification; its structure is shown in Fig. 2. The CNN model is implemented by Keras API running on top of Tensorflow using Python 3.7.

**GL with adaptive post-processing**

The adaptive decision boundary for group learning is determined using the statistical characteristics of training data for both classes. First, we take the mean value of the real number outputs of sub-images within each training sample (denoted as μi). Then we calculate the median number of μi for both positive and negative classes (denoted as μP, μN correspondingly). The adaptive decision boundary would be the average of μP and μN. A test sample would be classified as positive class, if and only if, its μt is greater than the decision boundary.

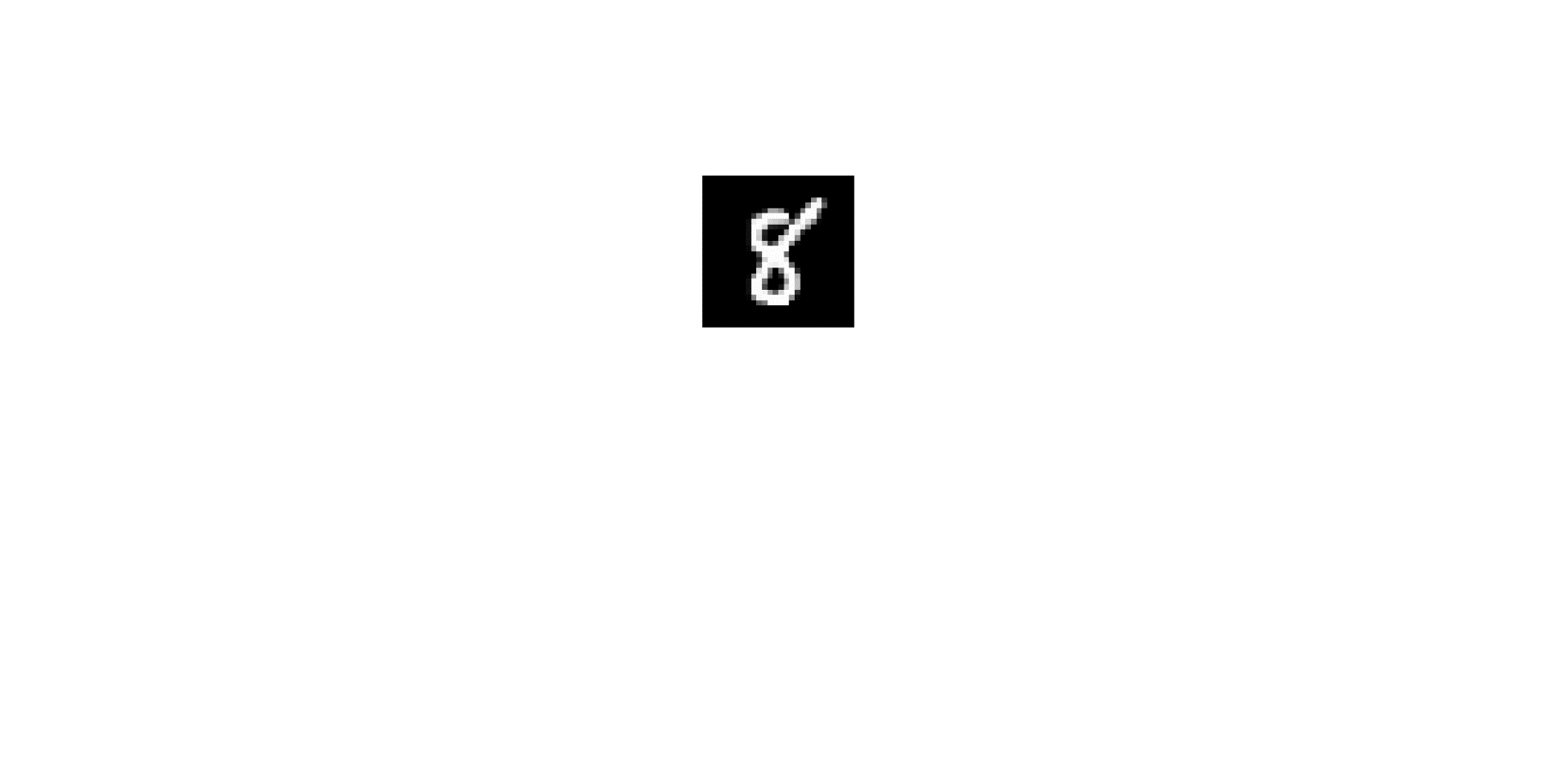


**Fig. 2.** CNN structure

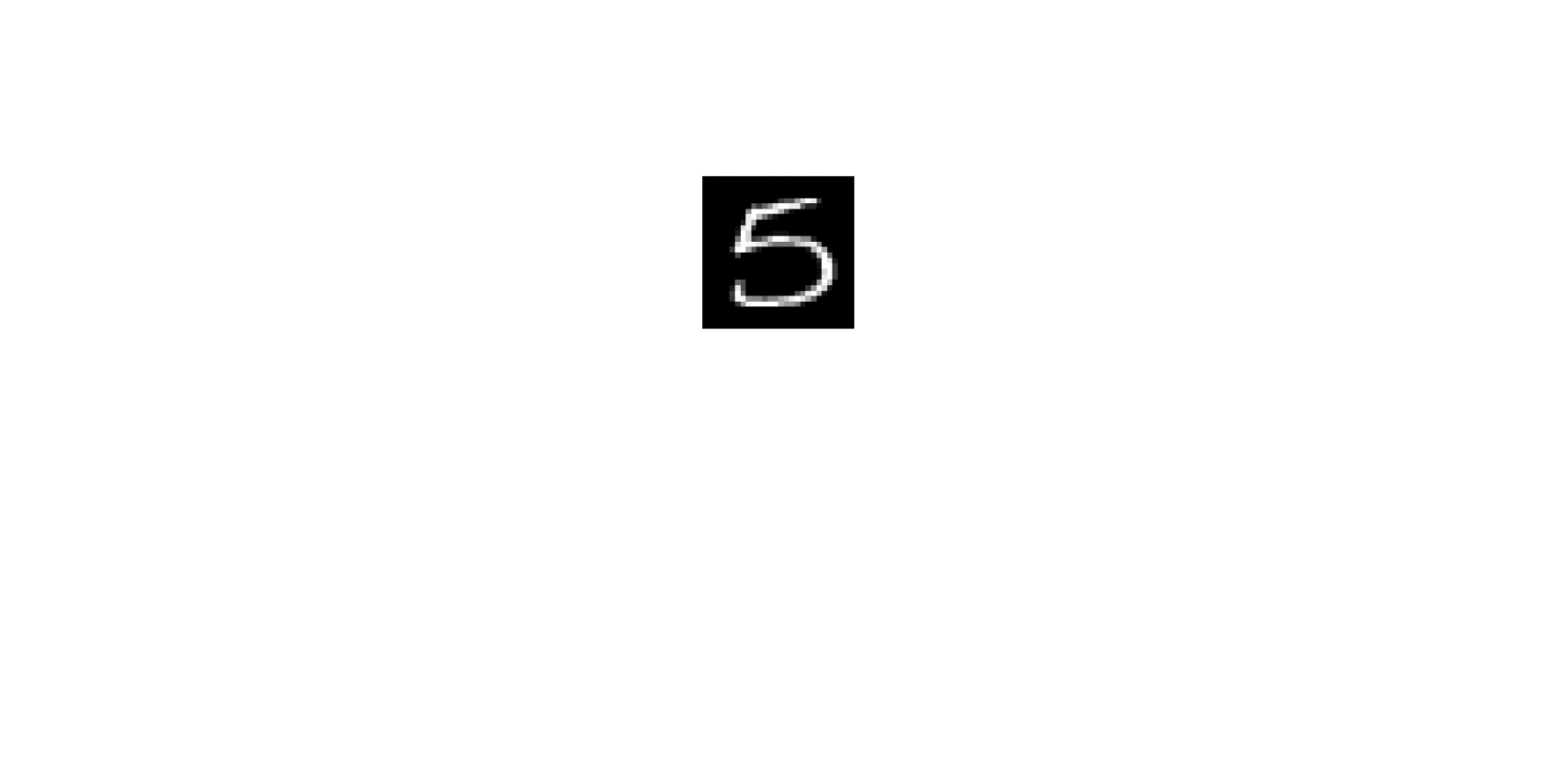
**Experiment list:**

* **Handwritten digit images:**
* **Digits 8 vs. 5**
* **Digits 3 vs. 5**
* **Handwritten letter images:**
* **letters e(E) vs. f(F)**
* **letters y(Y) vs. z(Z)**
* **Handwritten digit & letter images:**
* **Digits 7, 5 vs. letters j(J), r(R)**
* **Digits 6, 8 vs. letter n(N), s(S)**
* **Face images (yale dataset):**
* **Individuals 7 vs. 19**
* **Individuals 30 vs. 25**
* **Individuals 9 vs. 21**
* **Handwritten digit images**
* **Binary classification 1 (positive vs. negative samples):**

Positive samples (digit 8, 28\*28)

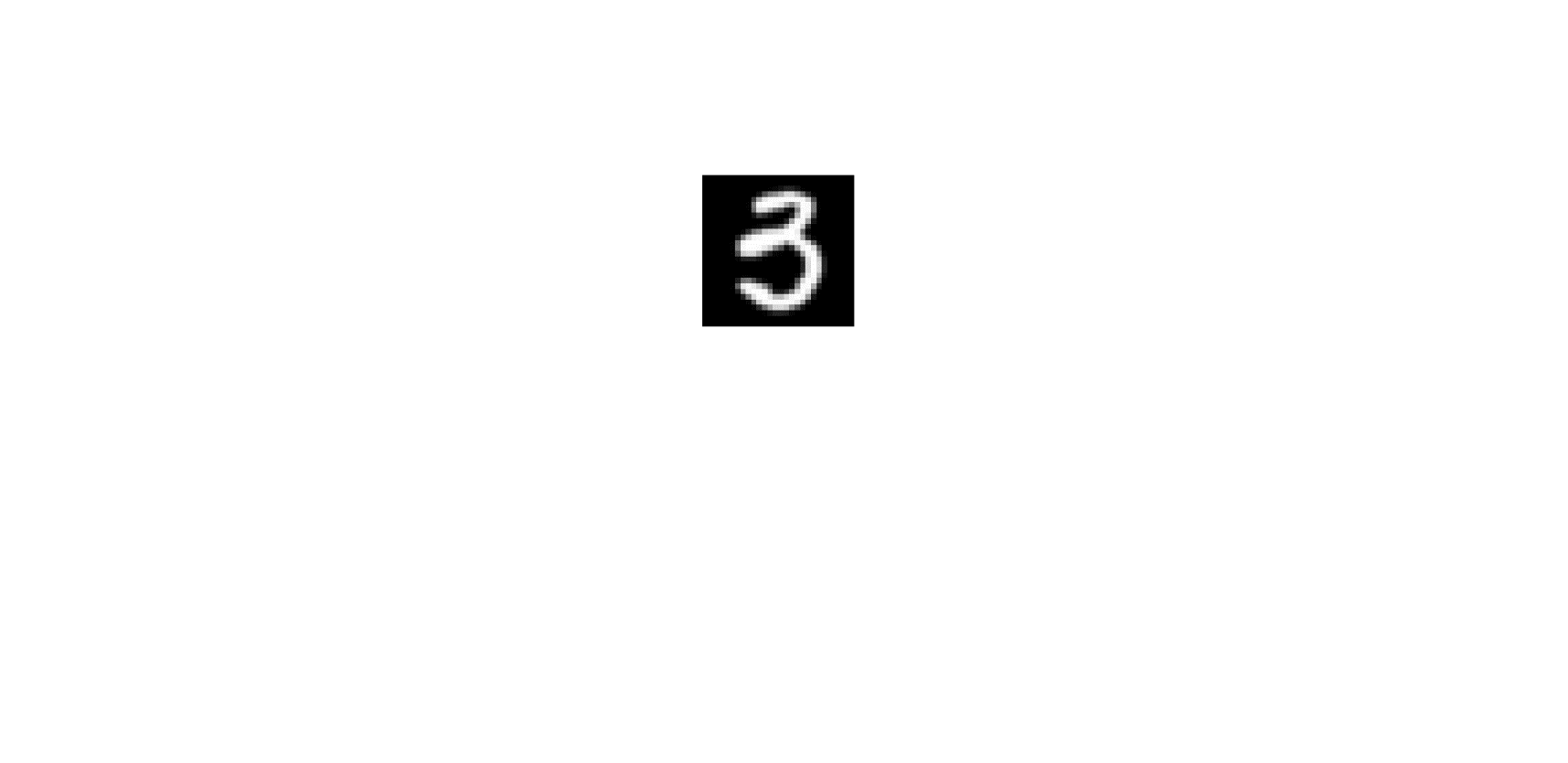


Negative samples (digit 5, 28\*28)

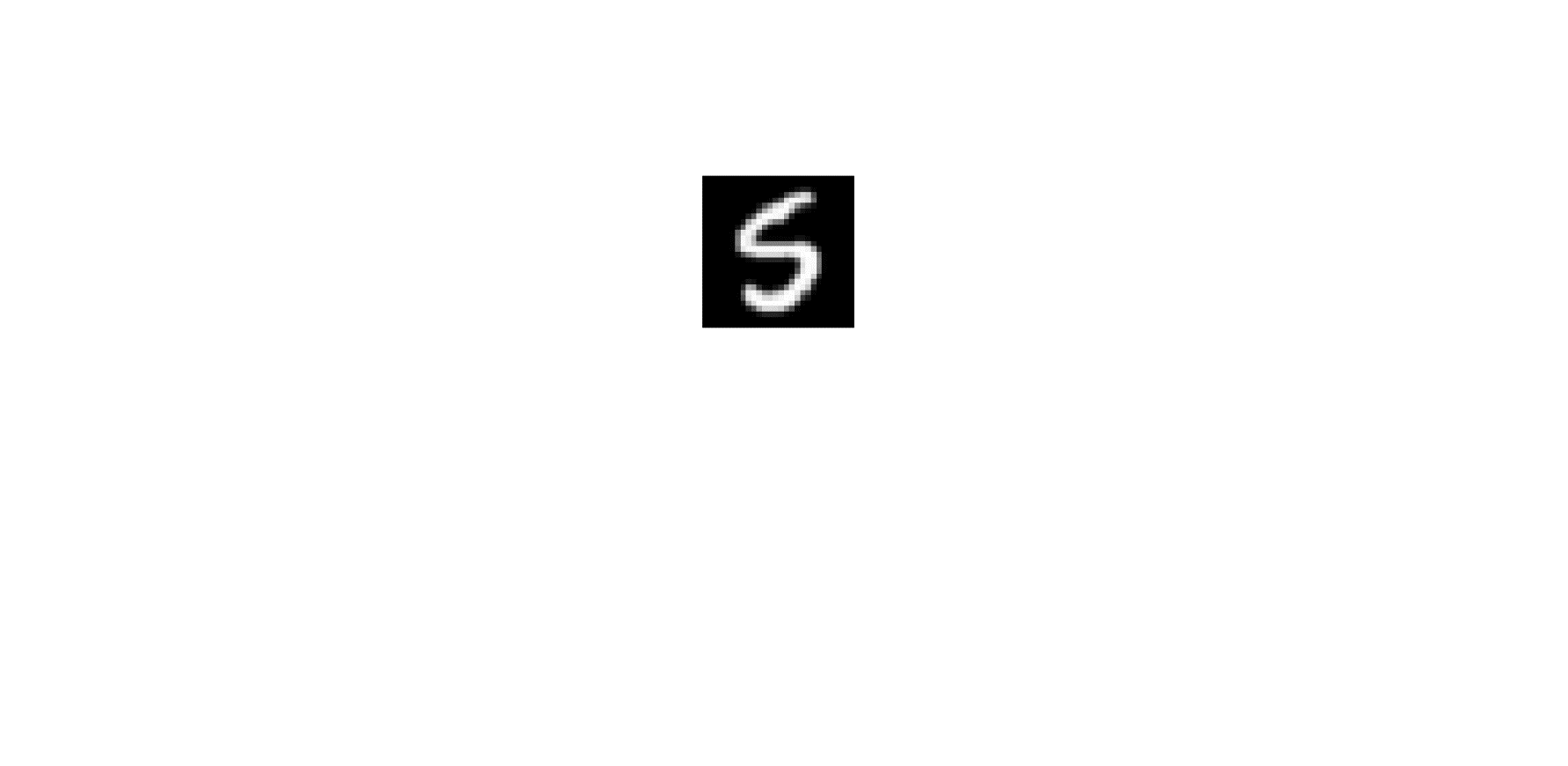


* **Binary classification 2 (positive vs. negative samples):**

Positive samples (digit 3, 28\*28)



Negative samples (digit 5, 28\*28)



* **Experimental setting:**

Training data: 10 (5 per class)

Validation data (for SVM): 10 (5 per class)

Test data: 1000 (500 per class)

All experiments are repeat 10 times.

* **Training/test errors for handwritten digit images:**

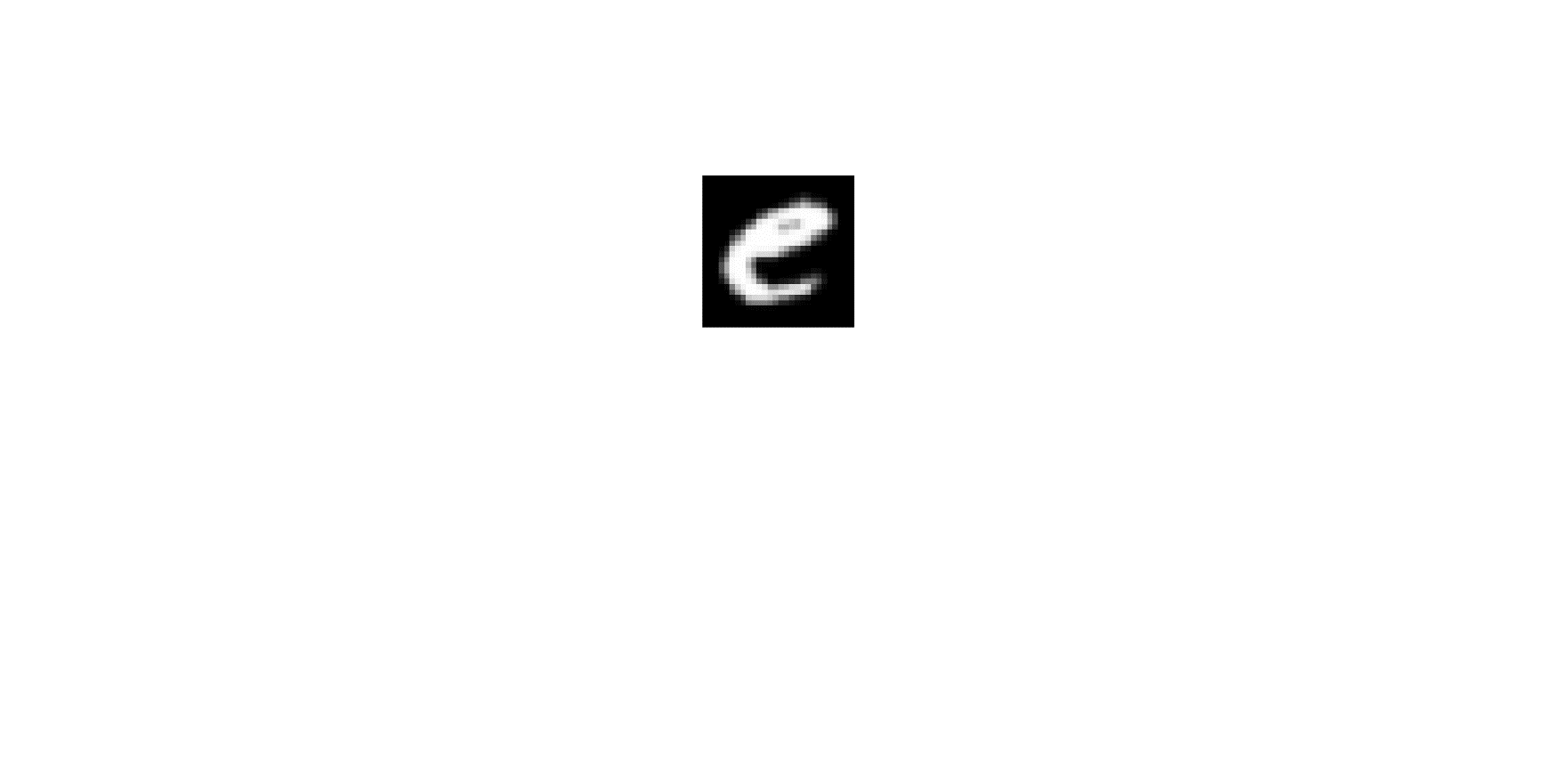
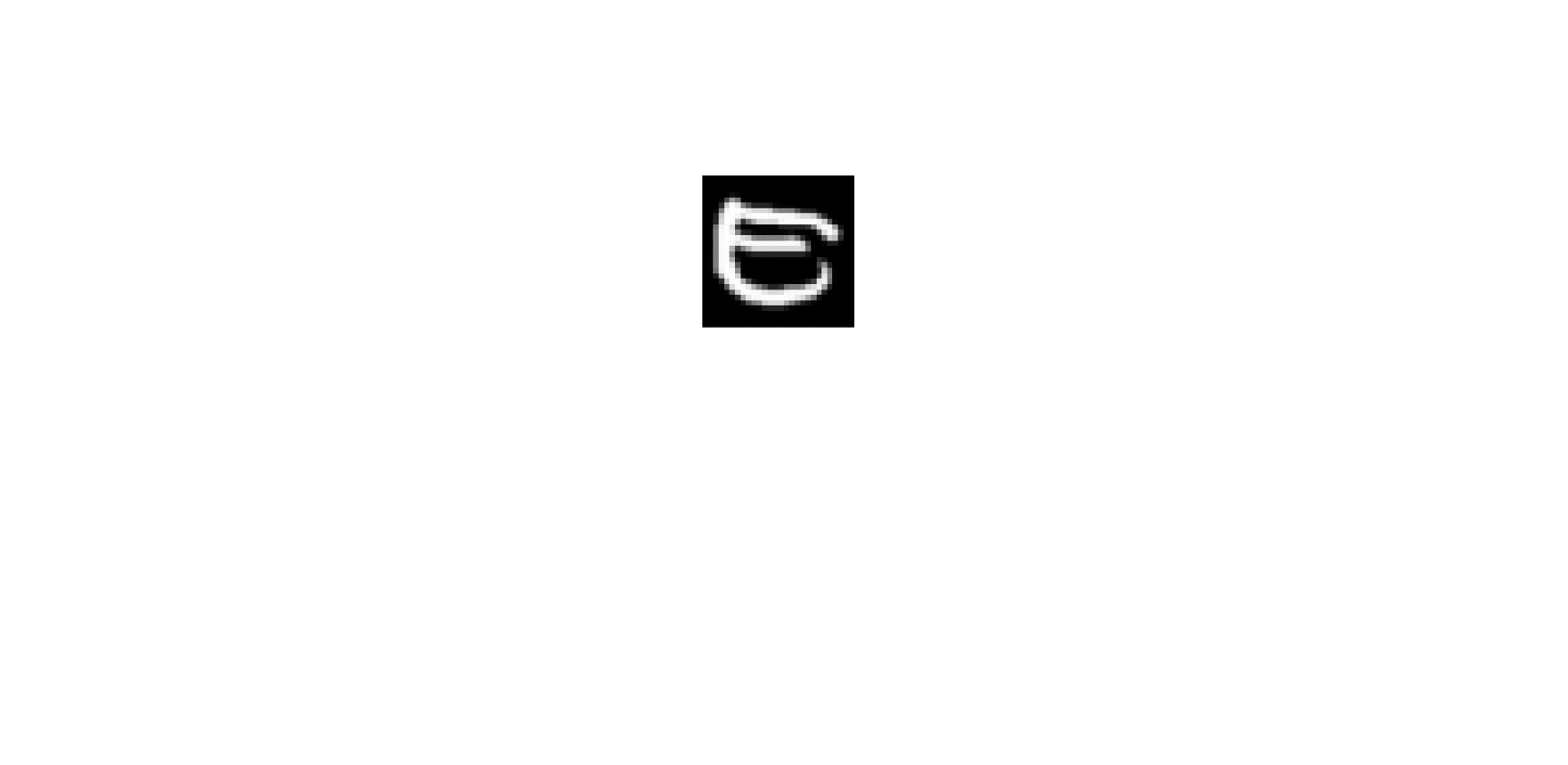
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **SVM** | | **SVM+GL(W=25)** | | **CNN** | | **CNN+GL(W=25)** | |
| **Training** | **Test** | **Training** | **Test** | **Training** | **Test** | **Training** | **Test** |
| **8 vs. 5** | 0 | 0.207 | 0 | 0.156 | 0 | 0.282 | 0 | 0.169 |
| **3 vs. 5** | 0 | 0.219 | 0 | 0.187 | 0 | 0.210 | 0 | 0.142 |

**Fig. 3.** **Digit 8 vs digit 5** results for SVM (left) and CNN (right). Horizontal line: standard method. Green bars: GL using majority voting method. Blue bars: GL using adaptive method.

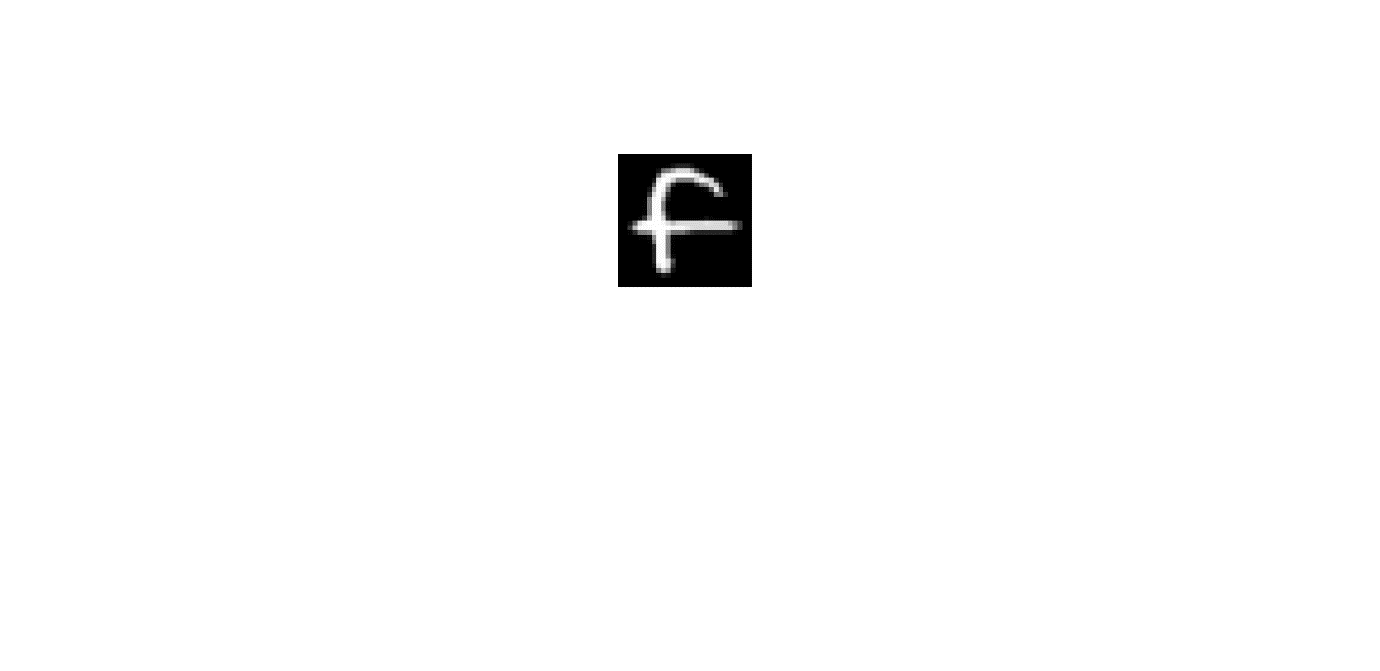
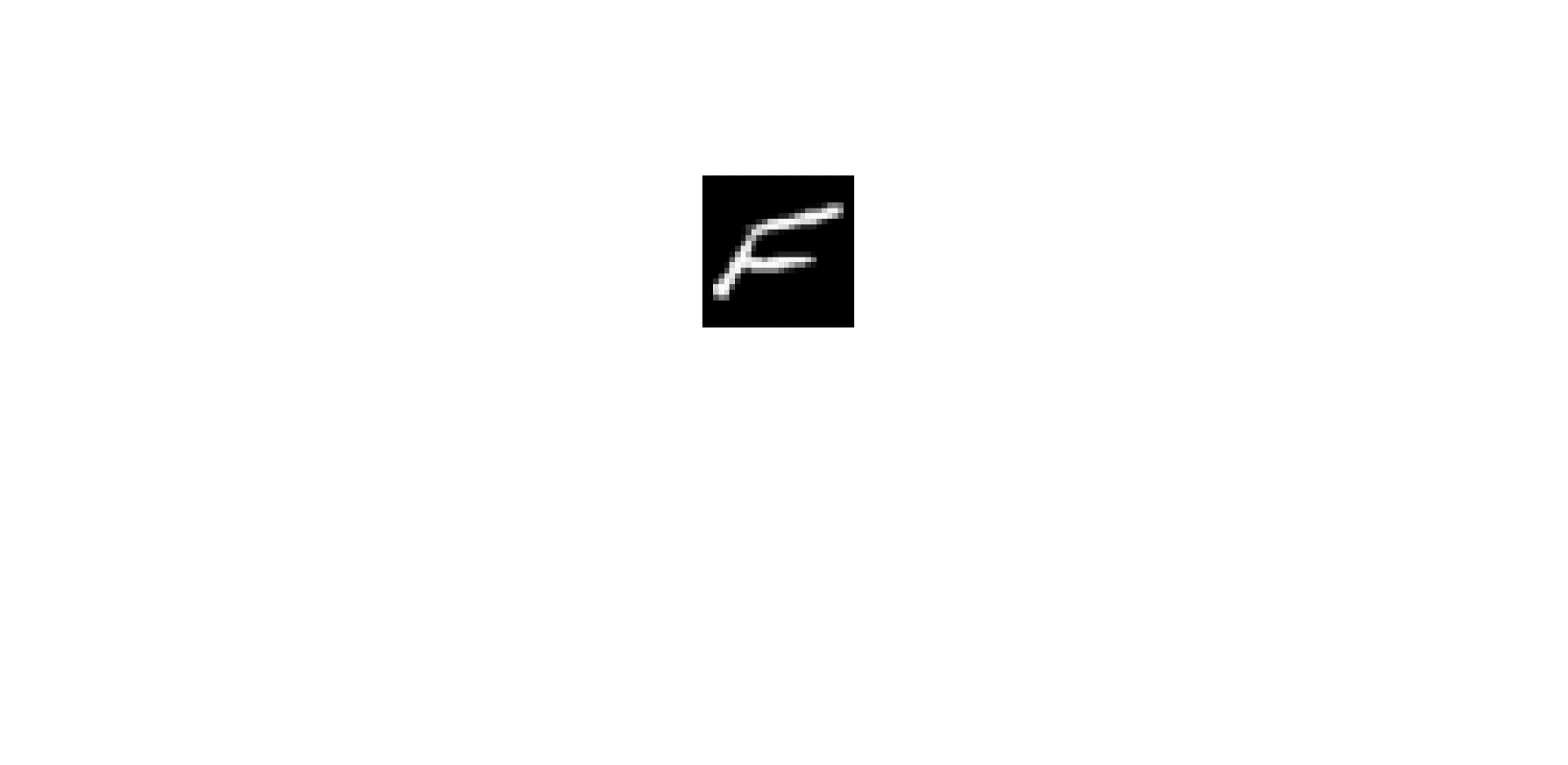
**Fig. 4.** **Digit 3 vs digit 5** results for SVM (left) and CNN (right). Horizontal line: standard method. Green bars: GL using majority voting method. Blue bars: GL using adaptive method.

* **Handwritten letter images**
* **Binary classification 3 (positive vs. negative samples):**

Positive samples (Letter e/E, 28\*28)

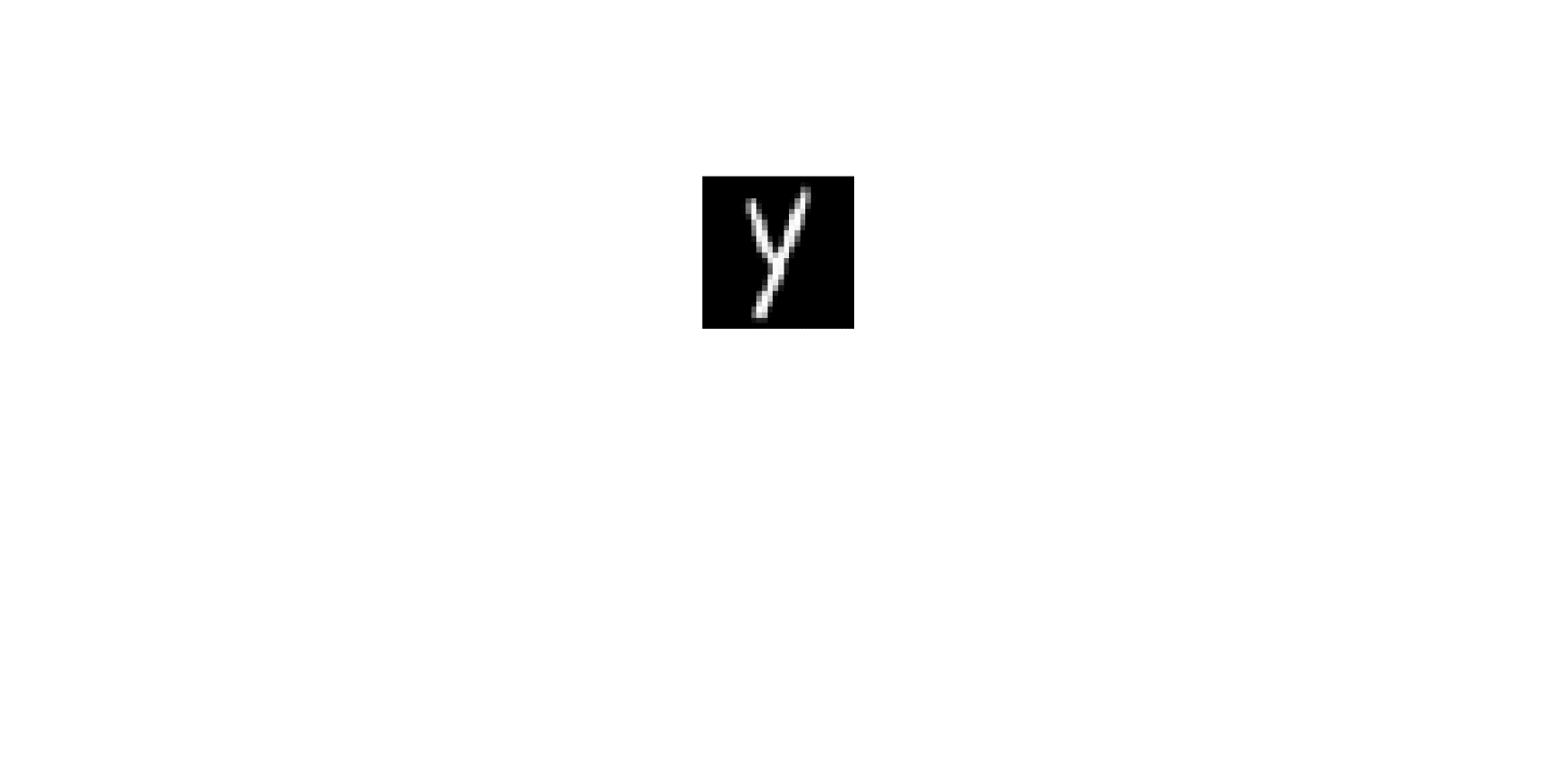
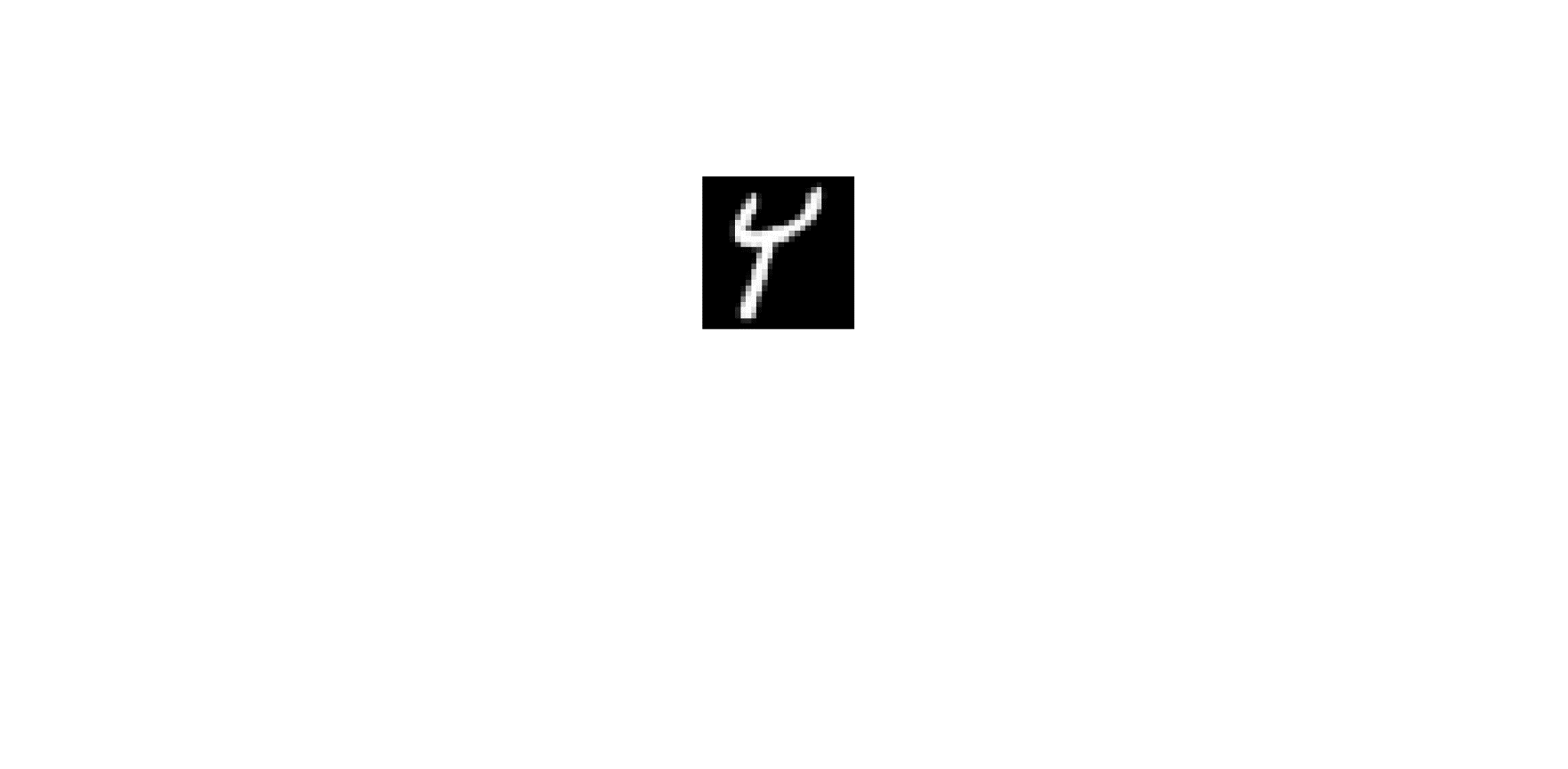
 

Negative samples (Letter f/F, 28\*28)

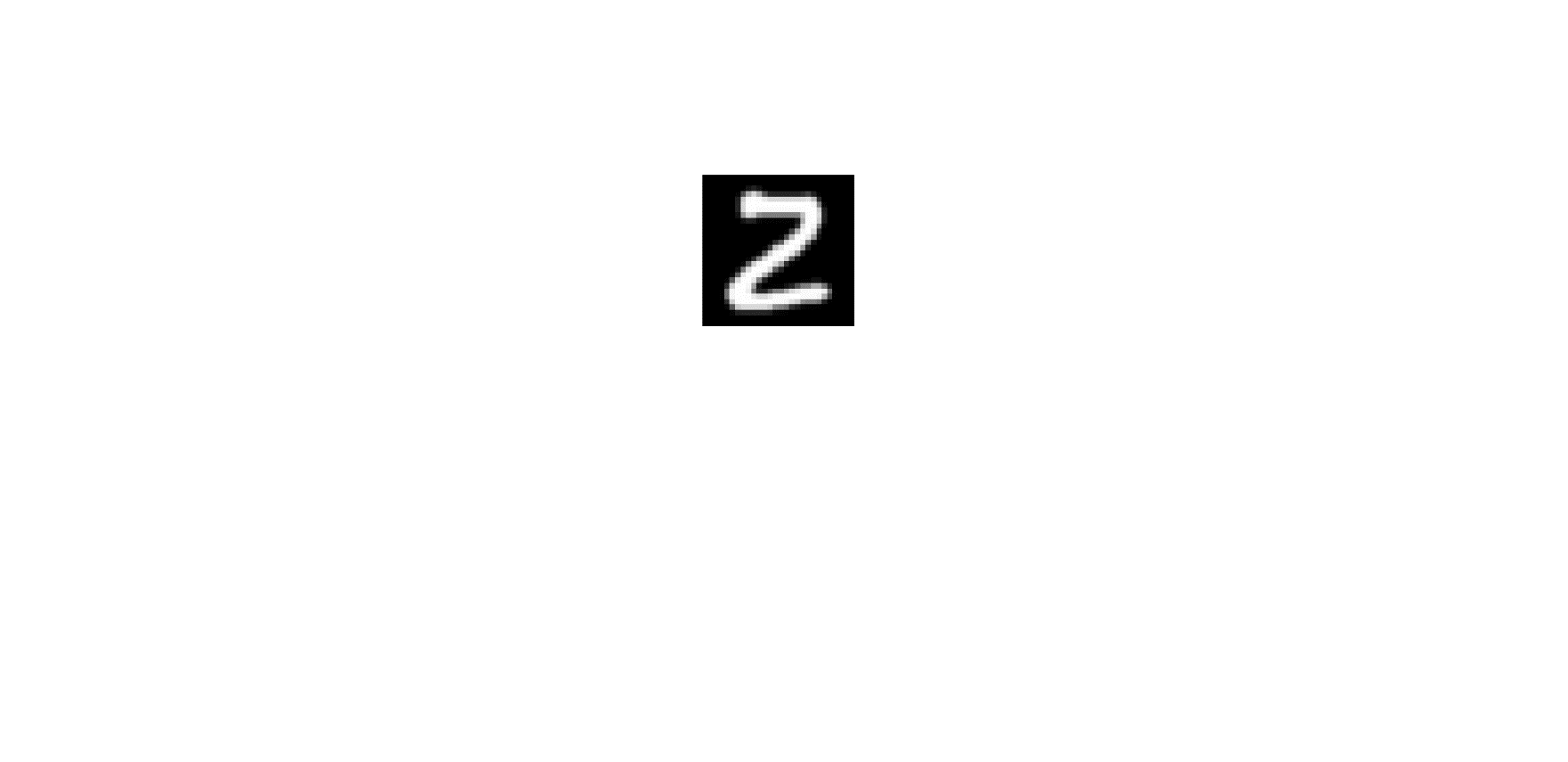
 

* **Binary classification 4 (positive vs. negative samples):**

Positive samples (Letter y/Y, 28\*28)

Negative samples (Letter z/Z, 28\*28)

* **Experimental setting:**

Training data: 10 (5 per class)

Validation data (for SVM): 10 (5 per class)

Test data: 1000 (500 per class)

All experiments are repeat 10 times.

* **Training/test errors for handwritten letter images:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **SVM** | | **SVM+GL(W=24)** | | **CNN** | | **CNN+GL(W=18)** | |
| **Training** | **Test** | **Training** | **Test** | **Training** | **Test** | **Training** | **Test** |
| **E vs. F** | 0 | 0.144 | 0 | 0.130 | 0.005 | 0.136 | 0 | 0.121 |
| **Y vs. Z** | 0 | 0.154 | 0 | 0.116 | 0 | 0.225 | 0 | 0.033 |

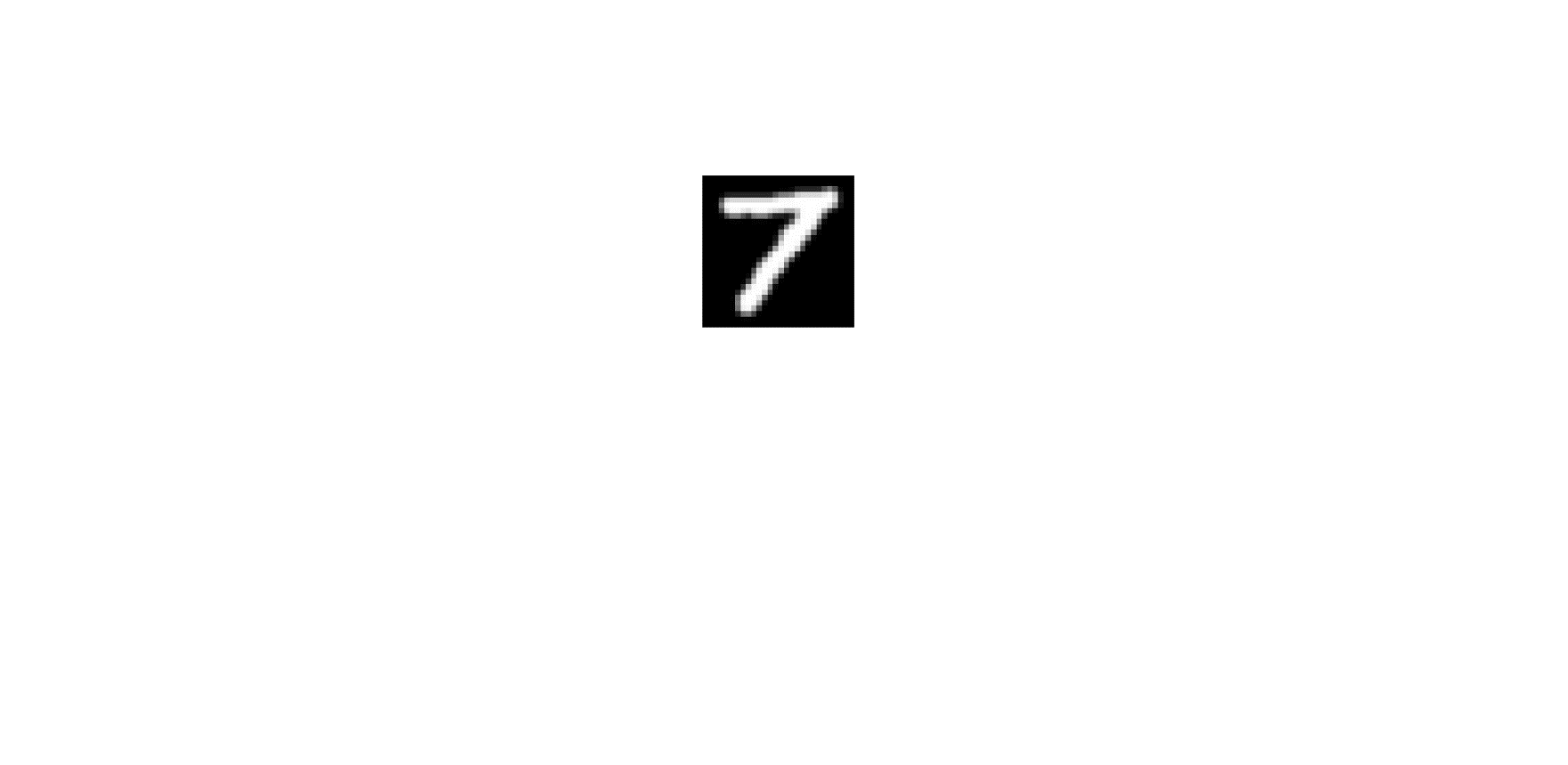
**Fig. 5.** **Letter E vs letter F** results for SVM (left) and CNN (right). Horizontal line: standard method. Blue bars: GL using adaptive method.

**Fig. 6.** **Letter Y vs letter Z** results for SVM (left) and CNN (right). Horizontal line: standard method. Blue bars: GL using adaptive method.

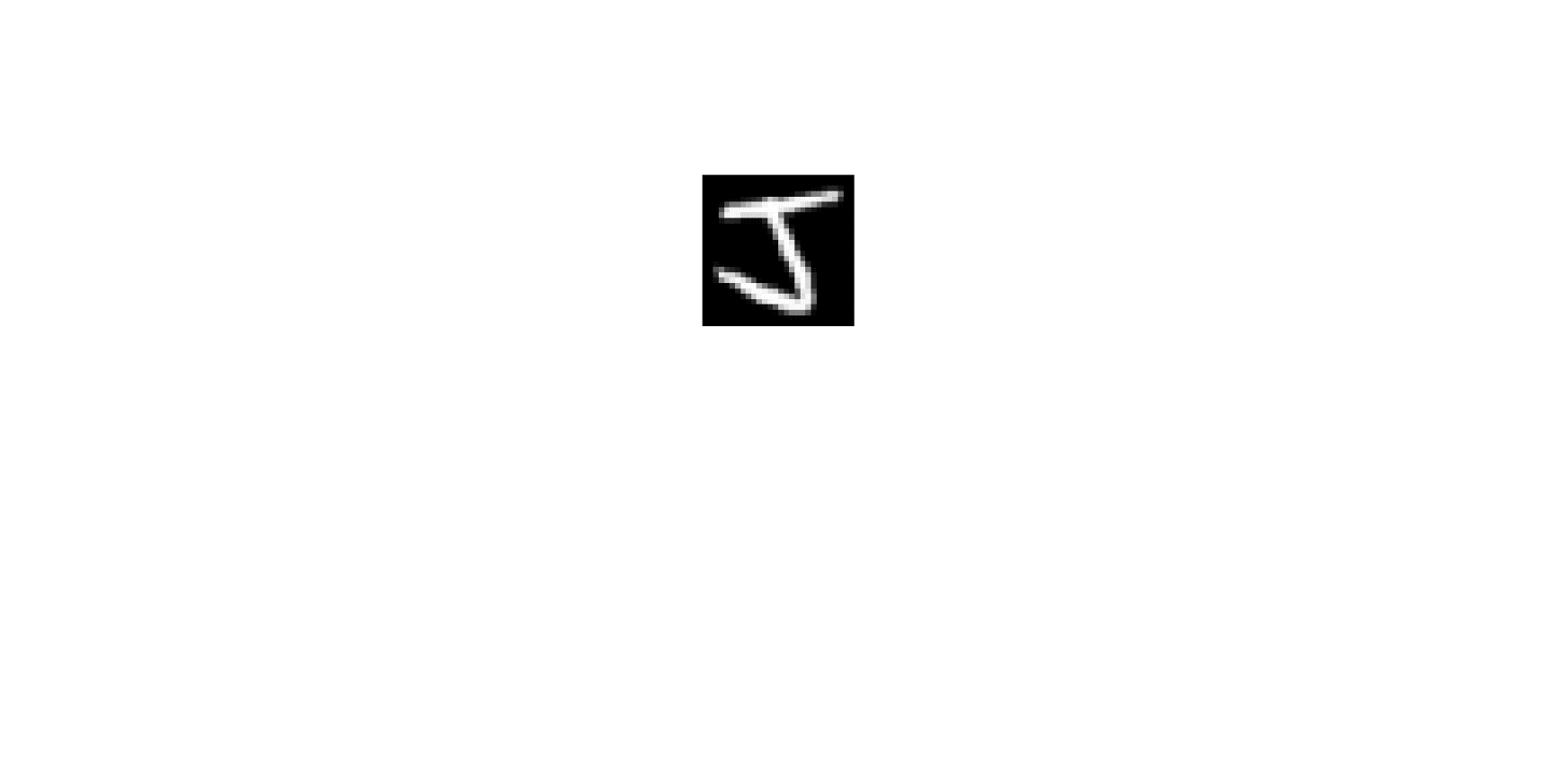
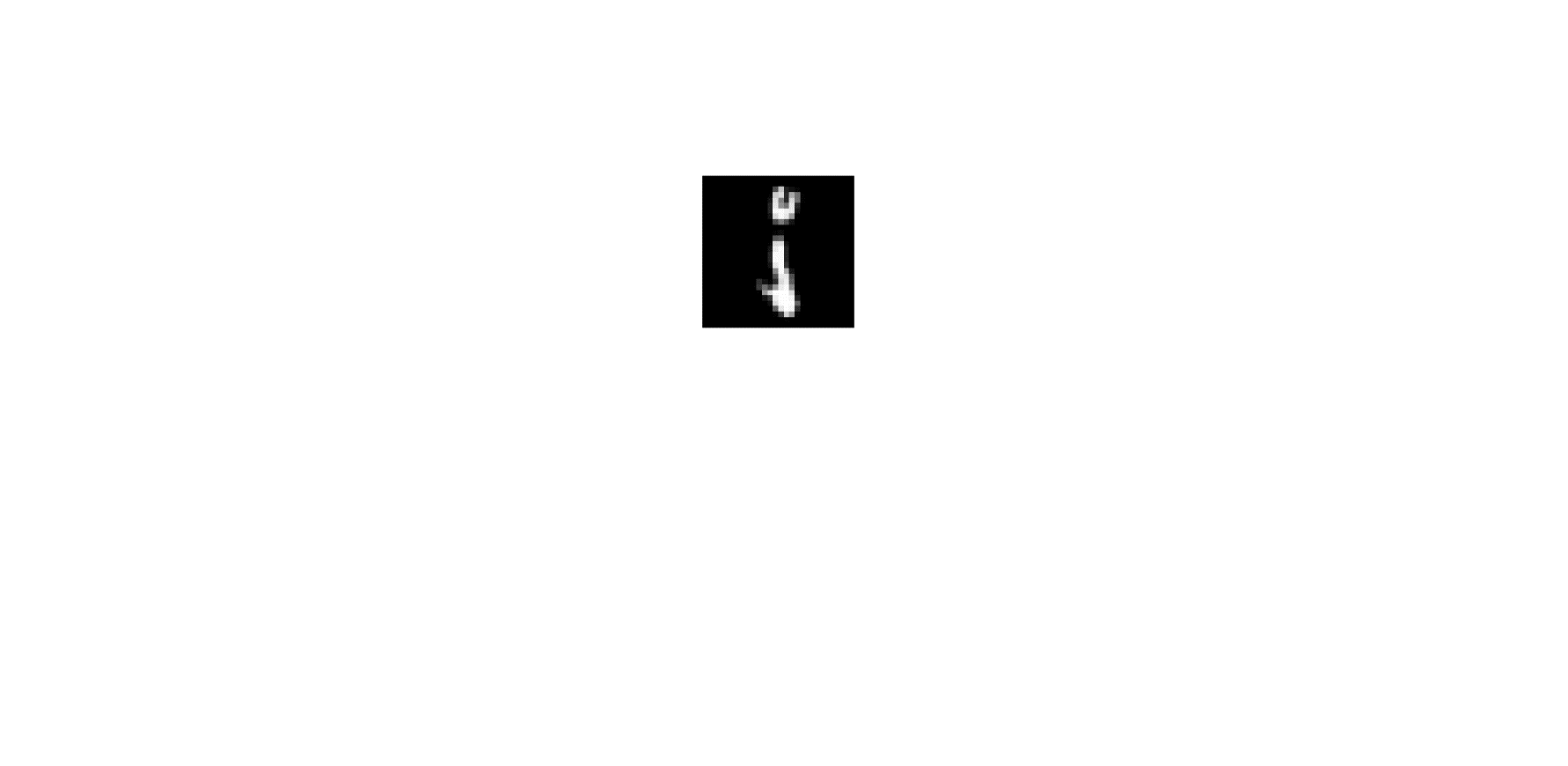
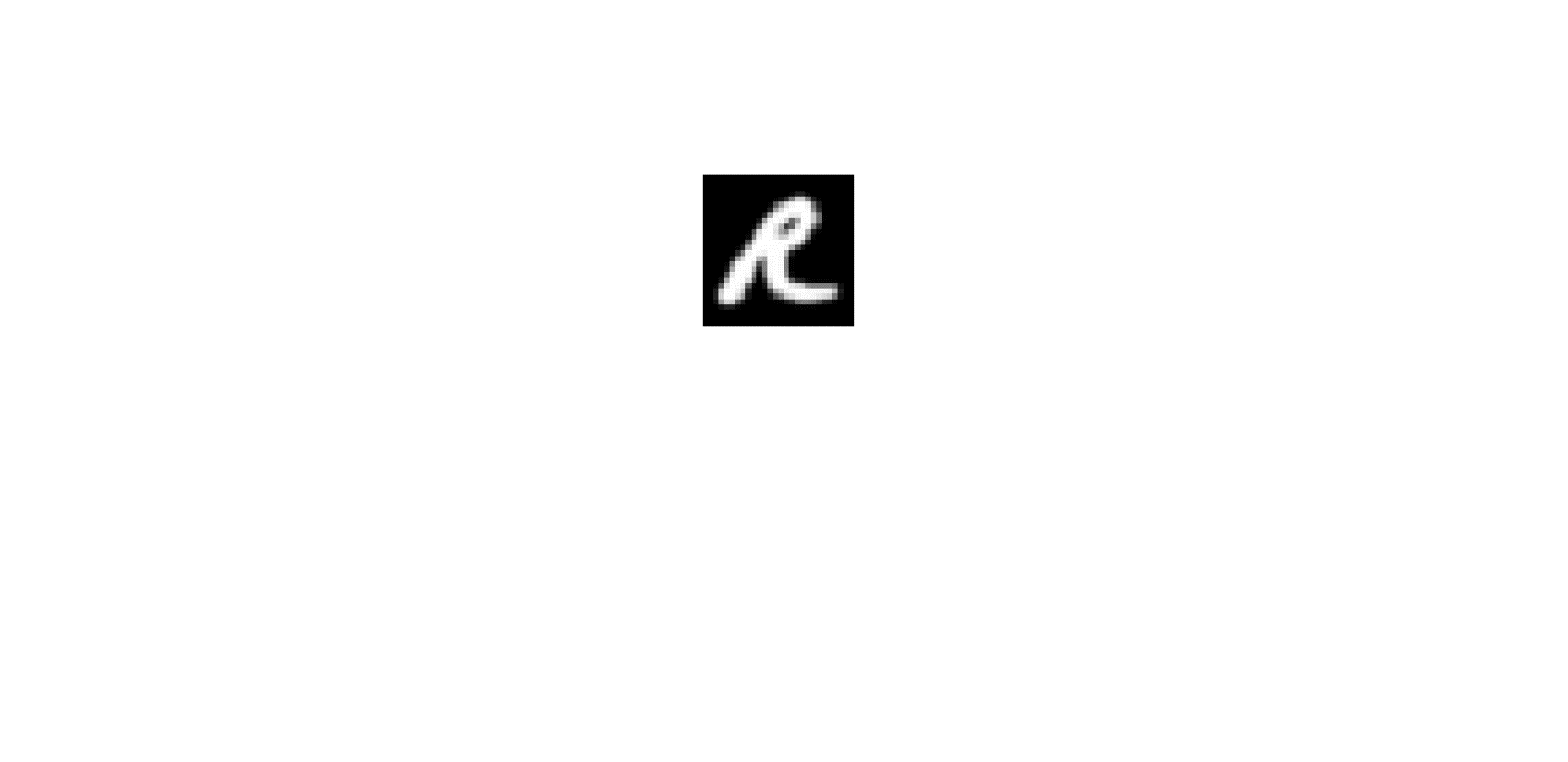
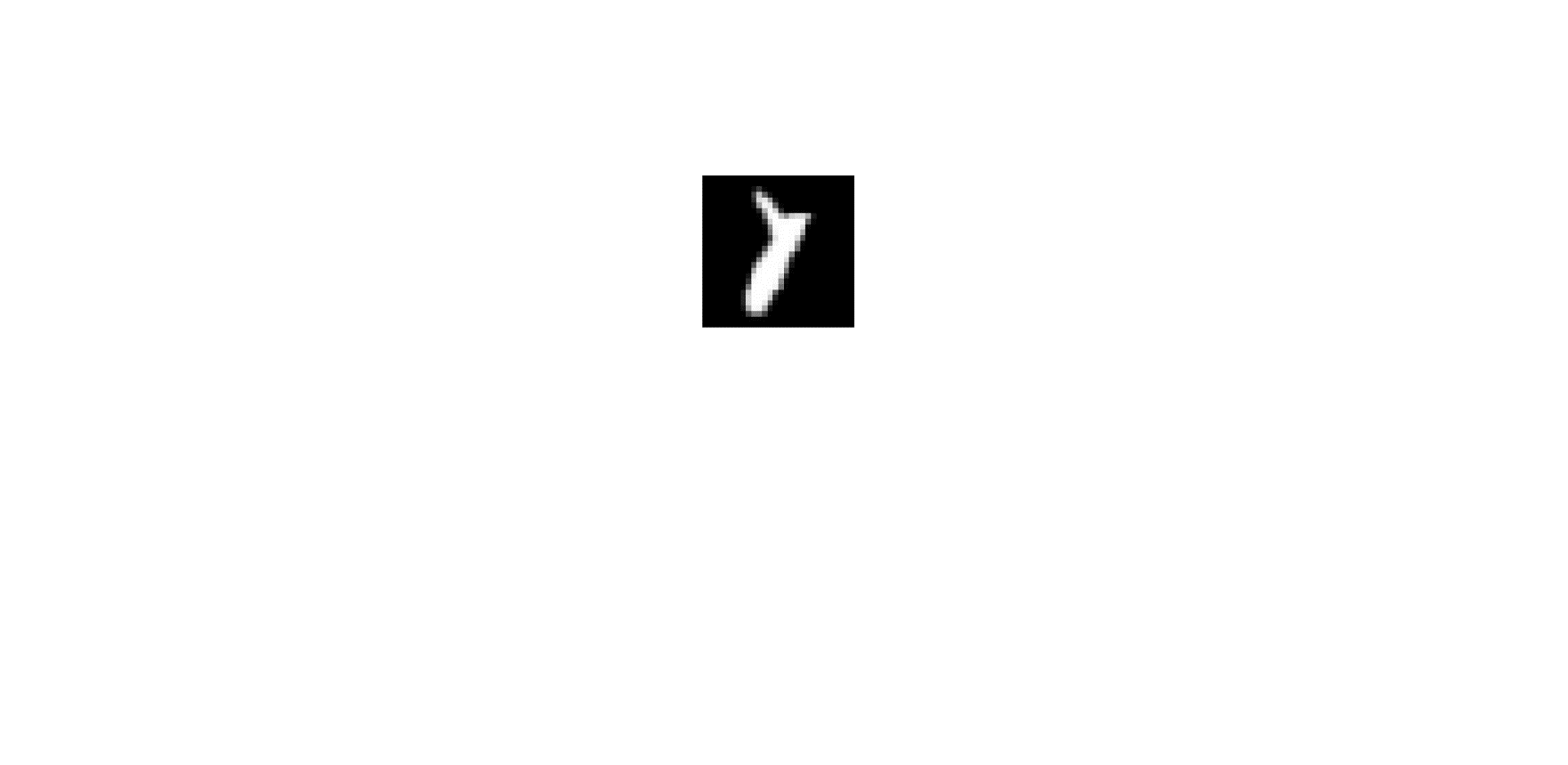
* **Handwritten digits/letters images**

**Binary classification 5 (positive vs. negative samples):**

Positive samples (randomly choose two digits, 7&5 from EMNIST dataset)

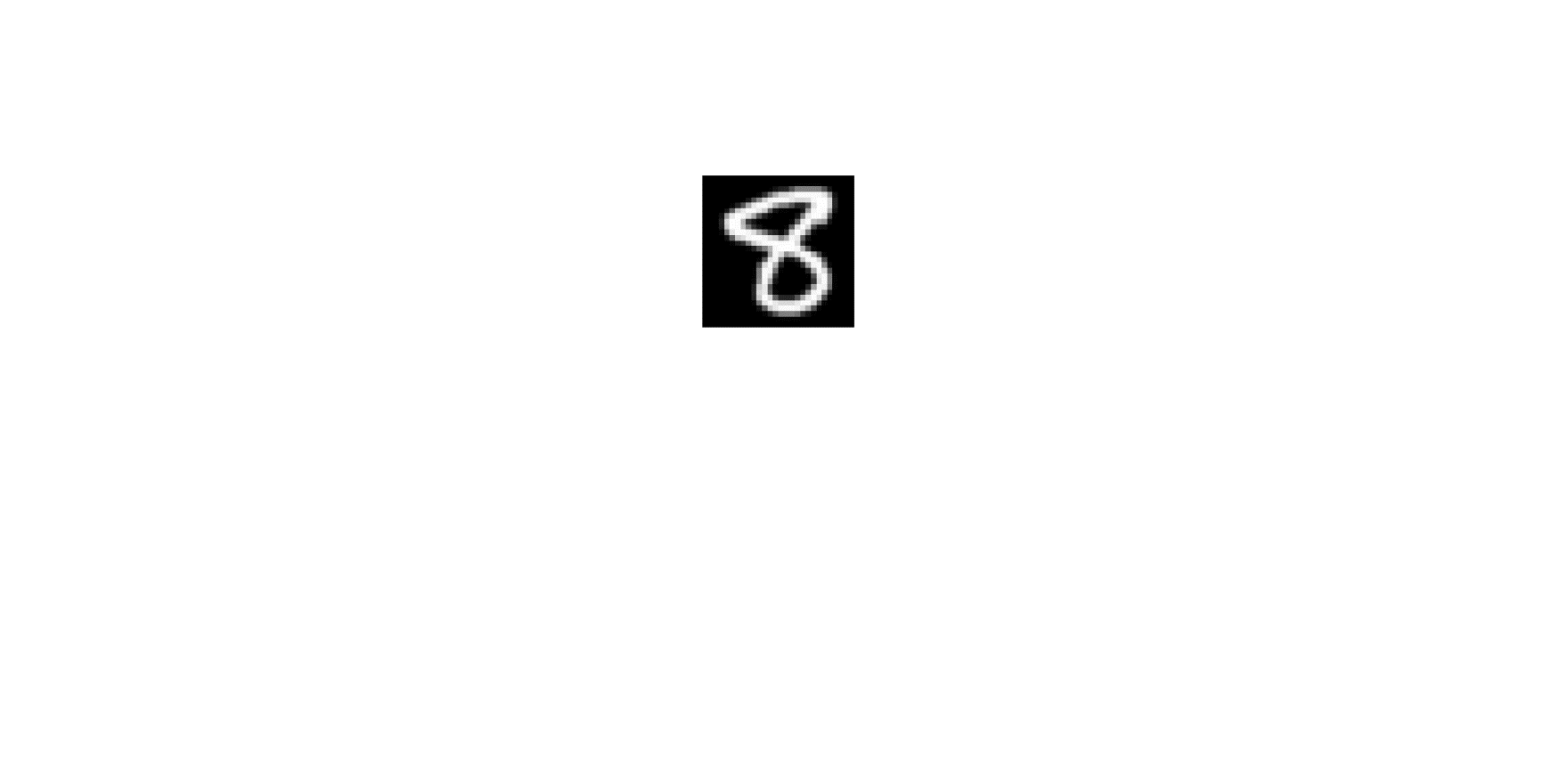
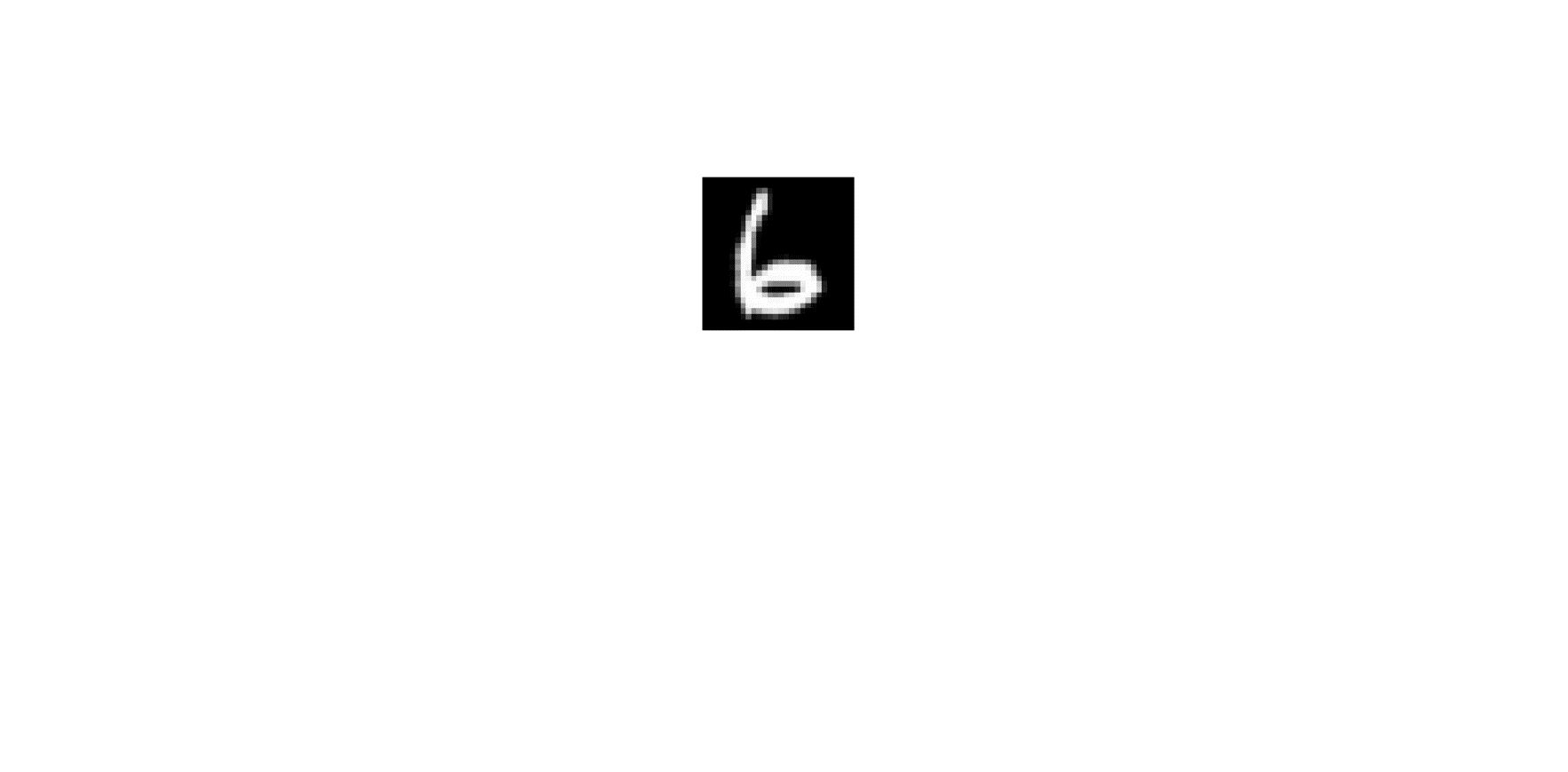
 

Negative samples (randomly choose two letters, J&R from EMNIST dataset)

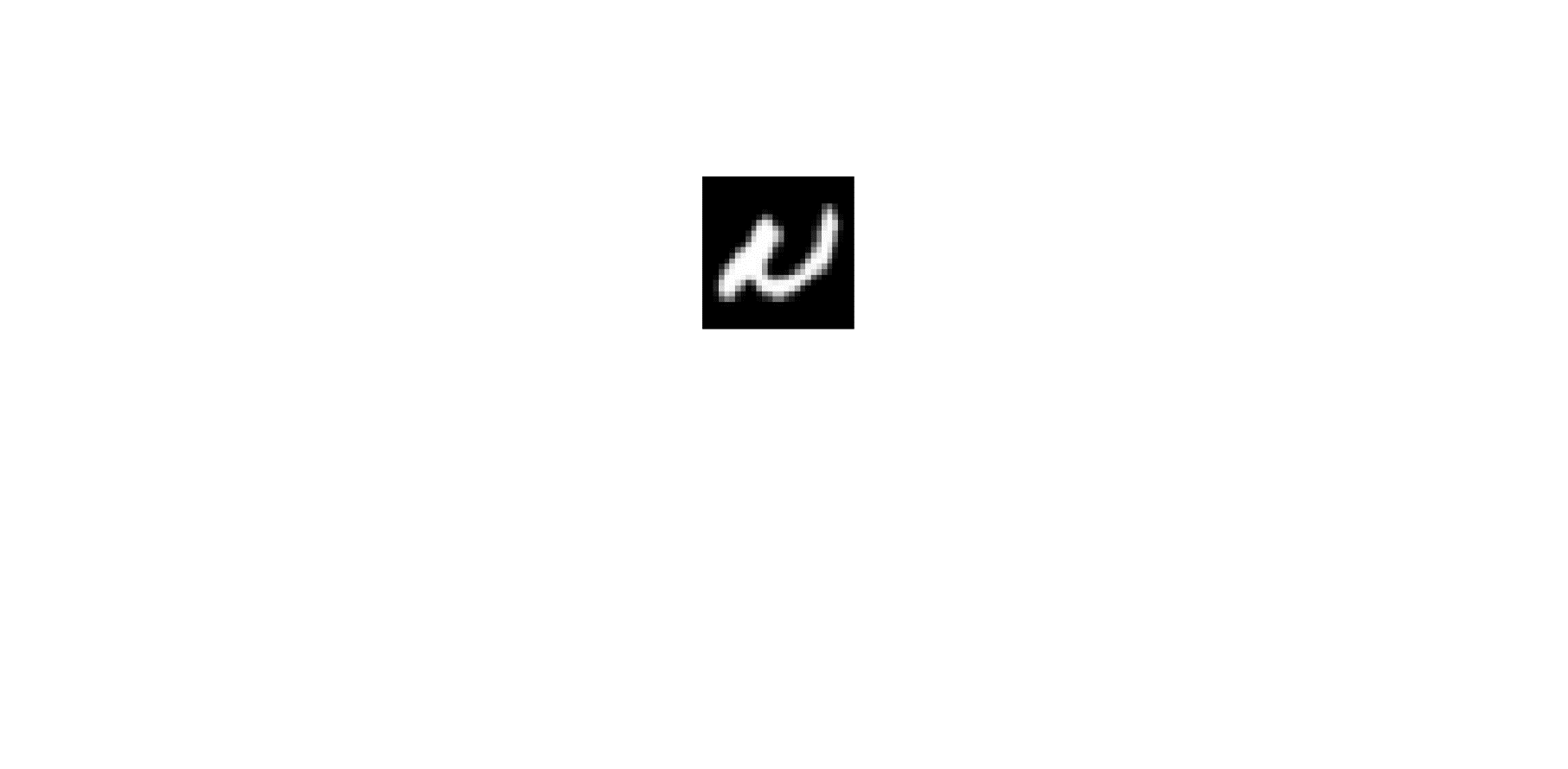
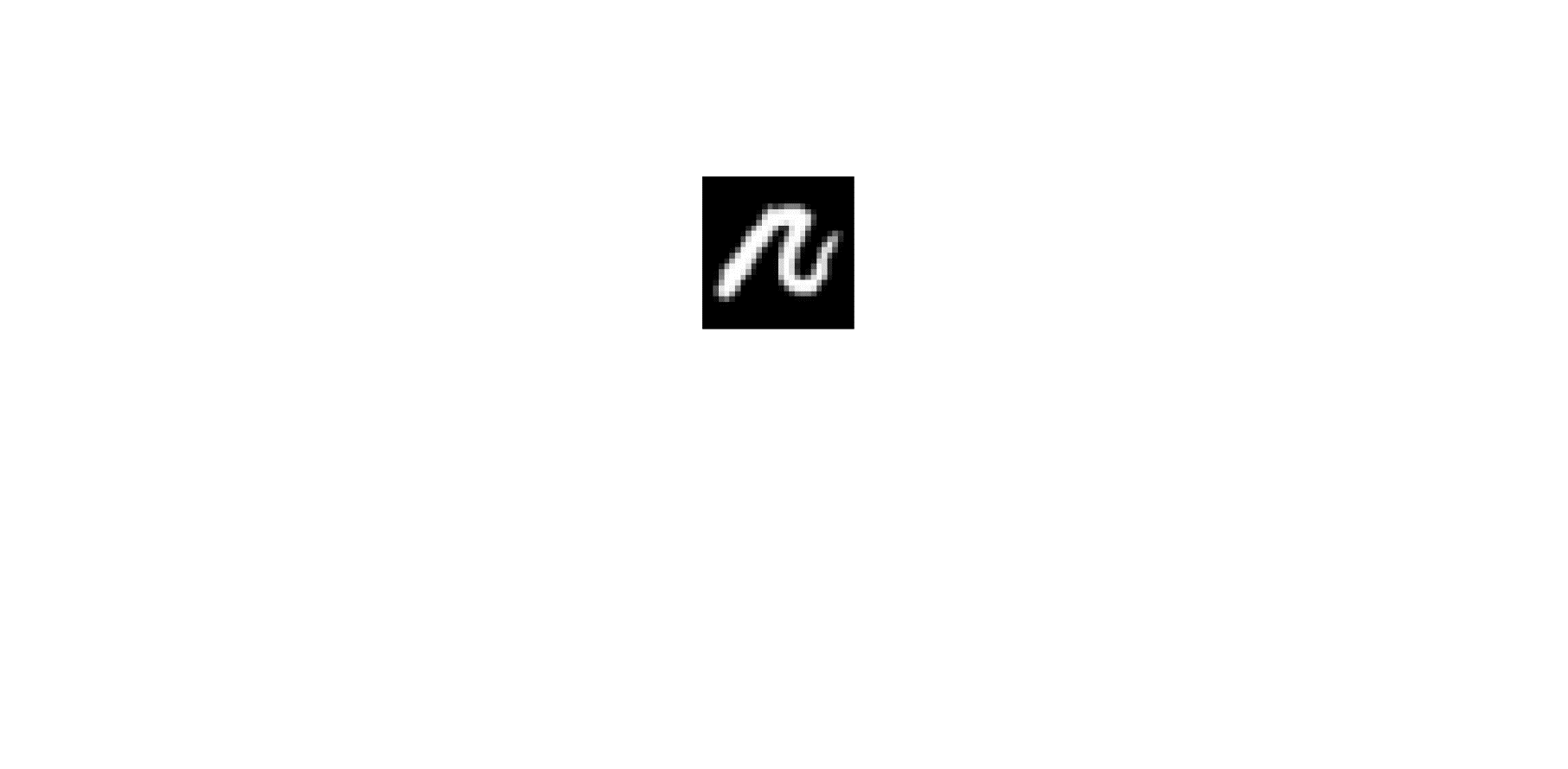
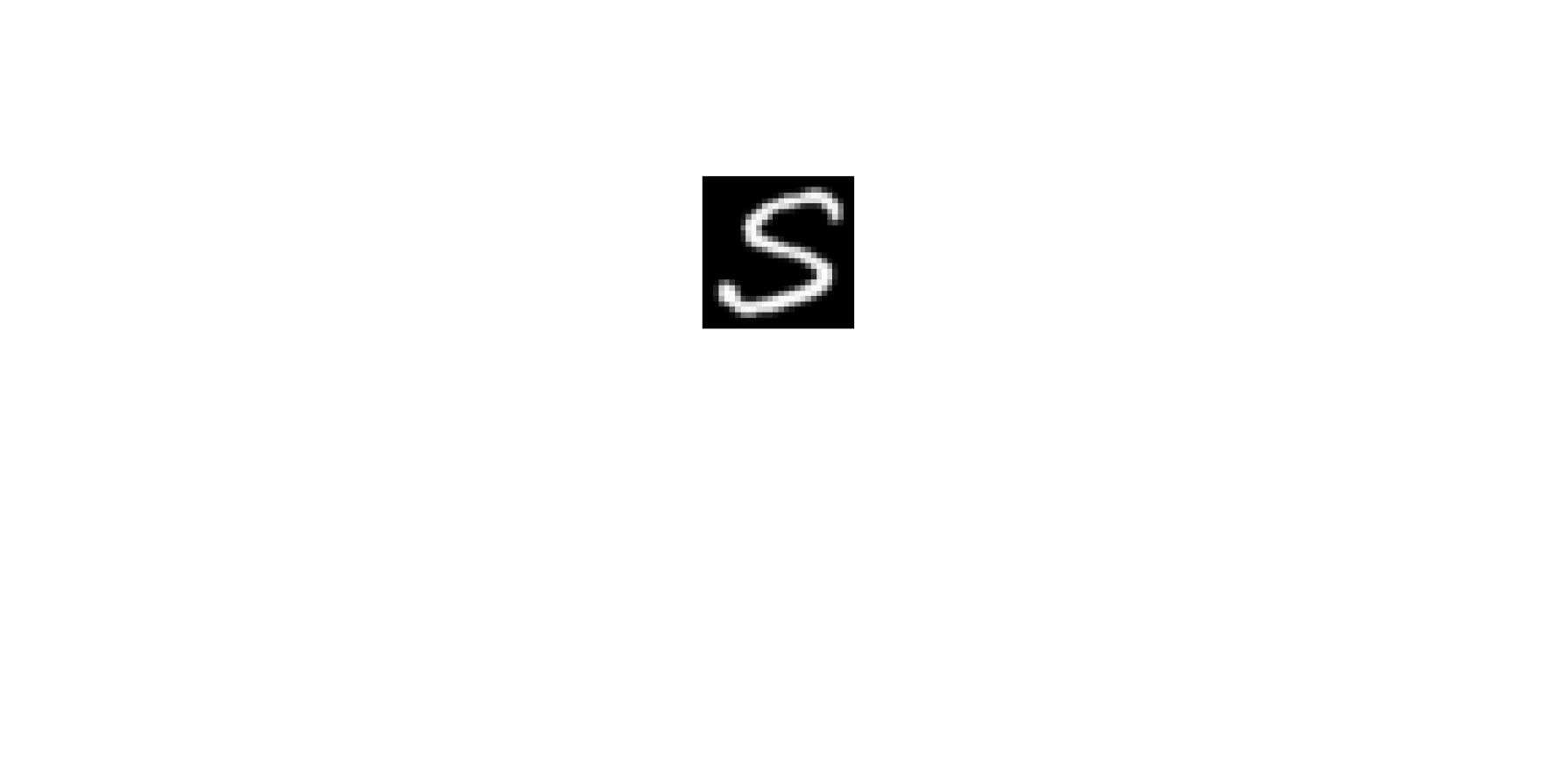
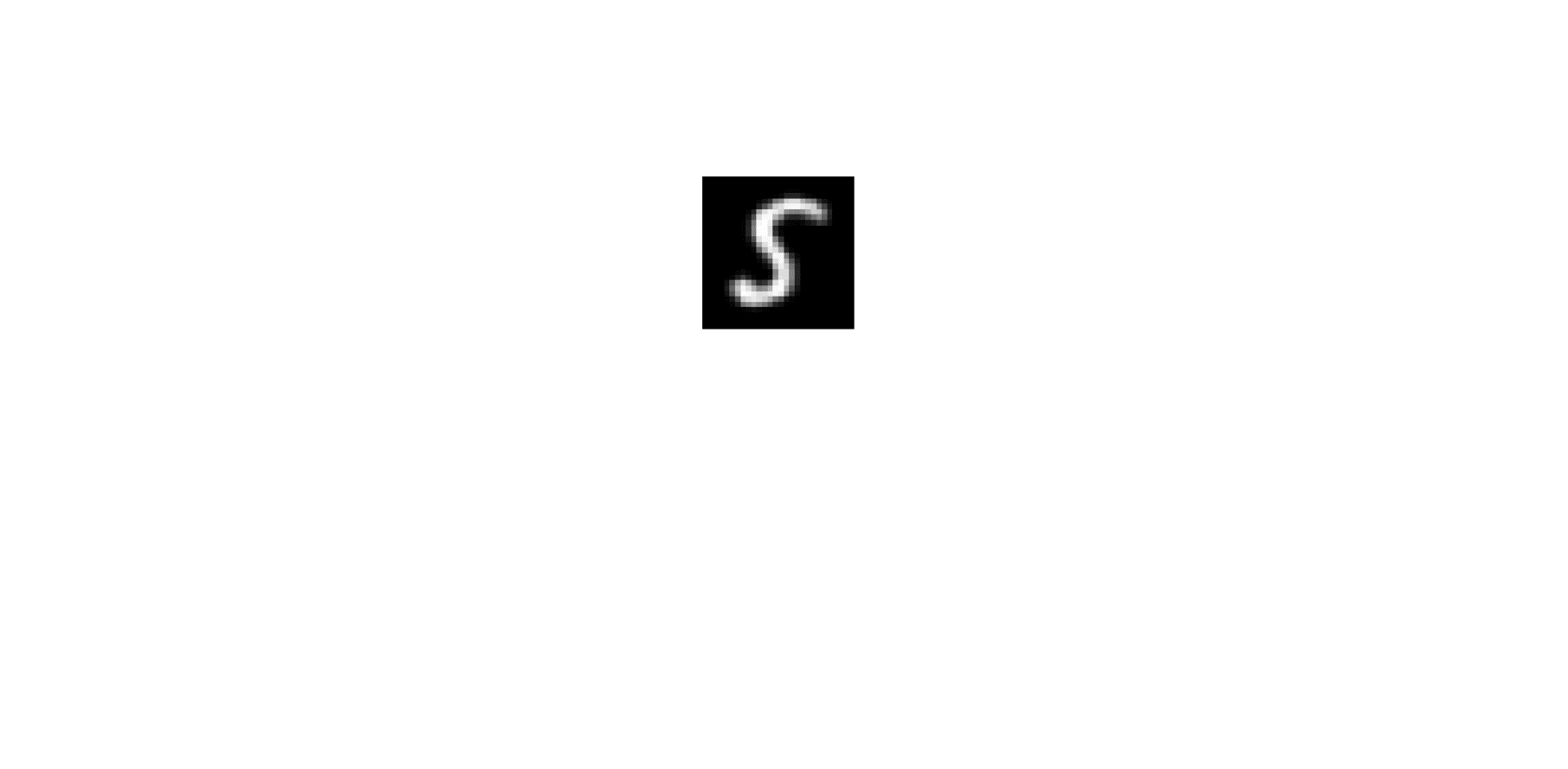
   

**Binary classification 6 (positive vs. negative samples):**

Positive samples (randomly choose two digits, 8&6 from EMNIST dataset)

Negative samples (randomly choose two letters, N&S from EMNIST dataset)

* **Experimental setting:**

Training data: 30 (15 per class)

Validation data (for SVM): 30 (15 per class)

Test data: 1000 (500 per class)

* **Training/test errors for handwritten letter images:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **SVM** | | **SVM+GL(W=18)** | | **CNN** | | **CNN+GL(W=21)** | |
| **Training** | **Test** | **Training** | **Test** | **Training** | **Test** | **Training** | **Test** |
| **7,5 vs. j,r** | 0 | 0.257 | 0 | 0.171 | 0.001 | 0.282 | 0.02 | 0.207 |
| **8,6 vs. n,s** | 0 | 0.178 | 0 | 0.112 | 0 | 0.186 | 0.041 | 0.152 |

**Fig. 7. Digits 5, 7 vs. Letters J, R** results for SVM (left) and CNN (right). Horizontal line: standard method. Blue bars: GL using adaptive method.

**Fig. 8. Digits 6, 8 vs. Letters N, S** results for SVM (left) and CNN (right). Horizontal line: standard method. Blue bars: GL using adaptive method.

* **Face images**

**Binary classification 7 (positive vs. negative samples):**

Positive samples (randomly choose face images of an individual, individual 7 is selected, from yale dataset, cropped to 160\*160)

一張含有 男人, 個人, 室內, 牆 的圖片

自動產生的描述 一張含有 男人, 個人, 眼鏡, 穿著 的圖片

自動產生的描述 一張含有 男人, 穿著, 凝視 的圖片

自動產生的描述

Negative samples (randomly choose face images of an individual, individual 19 is selected, from yale dataset, cropped to 160\*160)

一張含有 男人, 室內, 個人, 眼鏡 的圖片

自動產生的描述 一張含有 靈長類動物 的圖片

自動產生的描述 一張含有 眼罩, 耳, 靠近, 凝視 的圖片

自動產生的描述

**Binary classification 8 (positive vs. negative samples):**

Positive samples (randomly choose face images of an individual, individual 30 is selected, from yale dataset, cropped to 160\*160)

一張含有 室內, 穿著, 男人, 個人 的圖片

自動產生的描述 一張含有 男人, 個人, 穿著, 靈長類動物 的圖片

自動產生的描述 一張含有 穿著, 靈長類動物, 眼罩, 領帶 的圖片

自動產生的描述

Negative samples (randomly choose face images of an individual, individual 25 is selected, from yale dataset, cropped to 160\*160)

一張含有 室內, 個人, 男人, 牆 的圖片

自動產生的描述 一張含有 靈長類動物 的圖片

自動產生的描述 一張含有 靈長類動物, 凝視, 麥克風 的圖片

自動產生的描述

**Binary classification 9 (positive vs. negative samples):** individuals 9 vs 21

* **Experimental setting:**

Training data: 20 (10 per class)

Validation data (for SVM): 20 (10 per class)

Test data (all the rest): 88 (44 per class)

* **Training/test errors for handwritten letter images:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **SVM** | | **SVM+GL(W=60)** | | **CNN** | | **CNN+GL(W=60)** | |
| **Training** | **Test** | **Training** | **Test** | **Training** | **Test** | **Training** | **Test** |
| **F7 vs. F19** | 0 | 0.131 | 0 | 0.068 | 0.005 | 0.467 | 0.015 | 0.288 |
| **F30 vs. F25** | 0 | 0.129 | 0 | 0.075 | 0 | 0.508 | 0.012 | 0.390 |
| **F9 vs. F21** | 0 | 0.122 | 0.005 | 0.112 | 0.002 | 0.424 | 0.075 | 0.341 |

**Fig. 9. Face images of 7 vs. 19** Results for SVM (left) and CNN (right). Horizontal line: standard method. Blue bars: GL using adaptive method.

**Fig. 10.** **Face images of 30 vs. 25** results for SVM (left) and CNN (right). Horizontal line: standard method. Blue bars: GL using adaptive method.

**Fig. 11.** **Face images of 9 vs. 21** results for SVM (left) and CNN (right). Horizontal line: standard method. Blue bars: GL using adaptive method.