به نام خداوند بخشنده و مهربان

یادگیری عمیق

پروژه درس

محسن نقی‌پورفر ۹۴۱۰۶۷۵۷

۱ توضیح مختصری از کد

در این قسمت یک توضیح مختصر از کد زده شده داده می‌شود. کد این پروژه به صورت ماژولار پیاده‌سازی شده است که یک پکیج را می‌سازد. این پکیج به این صورت است از چند فایل کمکی مثل utils، dataset و ... کمک می گیرد. توابع کمکی در فایل utils و تمام اعمال مورد نیاز برای دیتاست در فایل dataset ذخیره شده است. همچنین در فایل‌های GAN، CGAN معماری های مورد نظر برای شبکه‌های مولد پیاده‌سازی شده است. برای این پکیج از Tensorflow استفاده شده است. همچنین کد موجود در فایل CNN، یک شبکه کانوولوشنی عمیق برای دسته‌بندی استفاده شده است.

۱.۱ توضیح فایل Dataset.py

در این فایل یک کلاس DataLoader پیاده‌سازی شده است که این کلاس برای اداره کردن همه توابع مختلف لازم برای داده‌های مورد نیاز برای شبکه‌های عصبی استفاده می‌شود.

۲.۱ توضیح فایل GAN.py

در این فایل یک شبکه‌عصبی GAN به وسیله خود پکیج تنسورفلو پیاده‌سازی شده است. این شبکه در قالب یک کلاس پیاده‌سازی شده است.

۳.۱ توضیح فایل CGAN.py

در این فایل، همانطور که از اسم آن نیز پیداست، شبکه‌عصبی CGAN به وسیله پکیج Keras پیاده‌سازی شده است. این پیاده‌سازی به صورت شی گرا بوده است. همچنین همه توابع لازم برای رسم نمودار، کشیدن کاراکترهای جدید و ذخیره فایل لاگ حاصل از یادگیری این شبکه نیز پیاده‌سازی شده است.

۴.۱ توضیح فایل CNN.py

در این فایل شبکه کانولوشنی عمیق برای تشخیص نوع کاراکتر ورودی برای داده حجیم داده شده استفاده شده است.

۲ توضیح معماری CNN

برای ساخت شبکه CNN برای دسته‌بندی حروف مختلف در بین فونت‌ها، از یک شبکه عصبی عمیقی استفاده شده است که دارای ۵ لایه کانوولوشن و Maxpooling که به معماری آن به صورت زیر بوده است. شکل مربوط به معماری این شبکه عصبی در زیر آمده است. برای عمیق‌تر کردن شبکه عصبی از ایده شبکه معروف VGG استفاده شده است و از فیلتر های کوچک استفاده شده است تا شبکه عمیق‌تر شود البته در لایه از فیلتر ۱۱ در ۱۱ استفاده شده است.

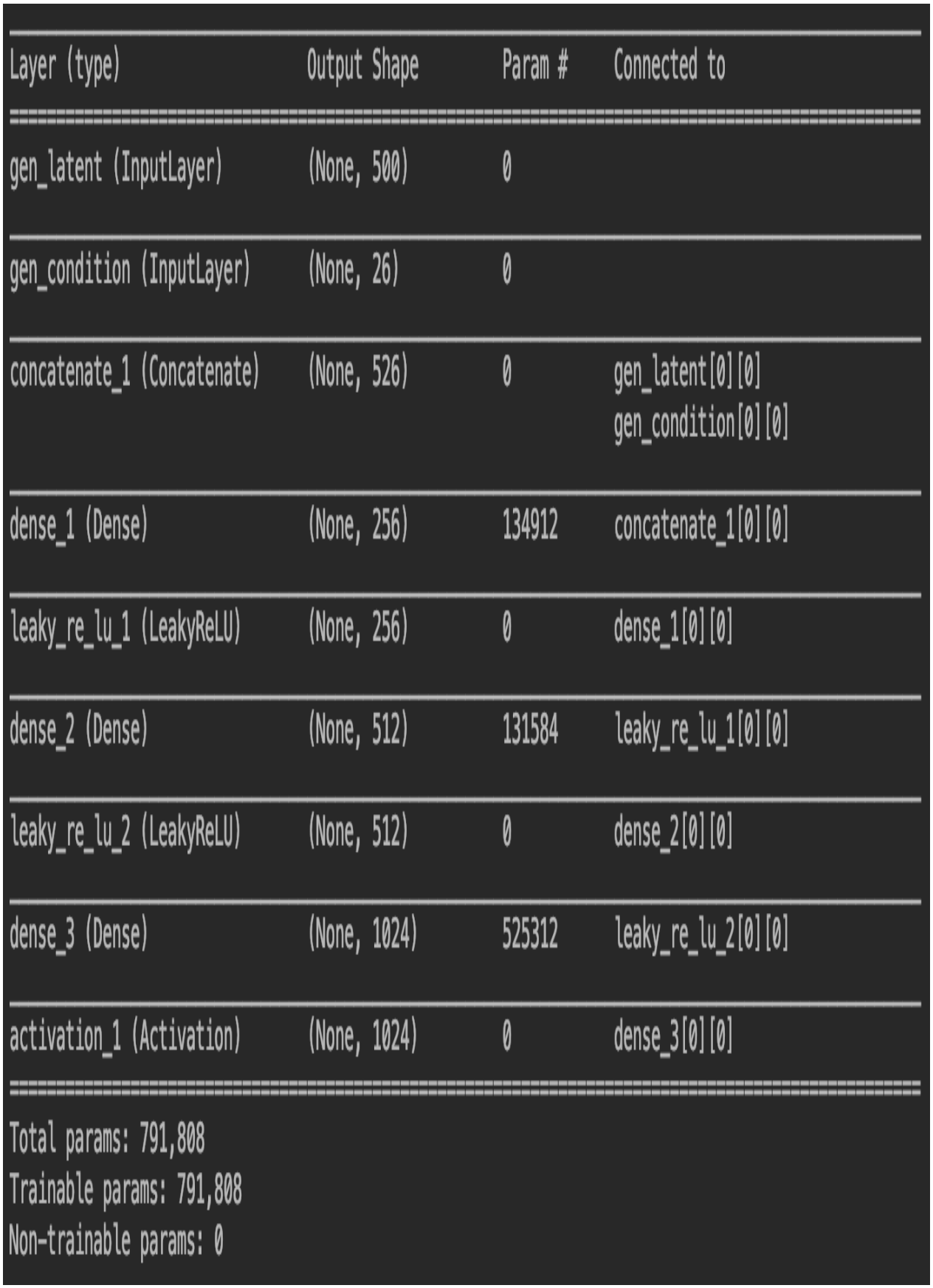
Layer (type)	Output Shape	Param #
inputs (InputLayer)	(None, 32, 32, 1)	0
conv_1 (Conv2D)	(None, 32, 32, 96)	11712
dropout_1 (Dropout)	(None, 32, 32, 96)	0
max_pooling2d_1 (MaxPooling2	(None, 16, 16, 96)	0
conv_2 (Conv2D)	(None, 16, 16, 256)	614656
dropout_2 (Dropout)	(None, 16, 16, 256)	0
max_pooling2d_2 (MaxPooling2	(None, 8, 8, 256)	0
conv_3 (Conv2D)	(None, 8, 8, 384)	885120
dropout_3 (Dropout)	(None, 8, 8, 384)	0
max_pooling2d_3 (MaxPooling2	(None, 4, 4, 384)	0
conv_4 (Conv2D)	(None, 4, 4, 384)	1327488
dropout_4 (Dropout)	(None, 4, 4, 384)	0
max_pooling2d_4 (MaxPooling2	(None, 2, 2, 384)	0
conv_5 (Conv2D)	(None, 2, 2, 384)	1327488
dropout_5 (Dropout)	(None, 2, 2, 384)	0
max_pooling2d_5 (MaxPooling2	(None, 1, 1, 384)	0
flatten_1 (Flatten)	(None, 384)	0
dense_1 (Dense)	(None, 512)	197120
dropout_6 (Dropout)	(None, 512)	0
outputs (Dense)	(None, 26)	13338

Total params: 4,376,922
Trainable params: 4,376,922
Non-trainable params: 0

شکل ۱: شبکه عصبی عمیق کانولوشنی استفاده شده در قسمت اول پروژه

۳ توضیح معماری Generator

برای شبکه Generator موجود در معماری‌های CGAN و GAN از دو معماری مختلف استفاده شده است. یکی شبکه‌ای کانولوشنی که از لایه Latent (به همراه لایه Condition در شبکه CGAN) و دیگری شبکه عصبی Fully Connected خالص که هر دو آزمایش شده‌اند و نتایج حاصل از آن‌ها موجود می‌باشد. برای شبکه CGAN،لایه Condition به لایه Latent چسبانده (Concatenate) شده است. شکل مربوط به معماری شبکه Generator در زیر آمده است.



Layer (type)	Output Shape	Param #	Connected to
gen_latent (InputLayer)	(None, 500)	0	
gen_condition (InputLayer)	(None, 26)	0	
concatenate_1 (Concatenate)	(None, 526)	0	gen_latent[0][0] gen_condition[0][0]
dense_1 (Dense)	(None, 256)	134912	concatenate_1[0][0]
leaky_re_lu_1 (LeakyReLU)	(None, 256)	0	dense_1[0][0]
dense_2 (Dense)	(None, 512)	131584	leaky_re_lu_1[0][0]
leaky_re_lu_2 (LeakyReLU)	(None, 512)	0	dense_2[0][0]
dense_3 (Dense)	(None, 1024)	525312	leaky_re_lu_2[0][0]
activation_1 (Activation)	(None, 1024)	0	dense_3[0][0]

Total params: 791,808
Trainable params: 791,808
Non-trainable params: 0

شکل ۲: معماری شبکه Generator

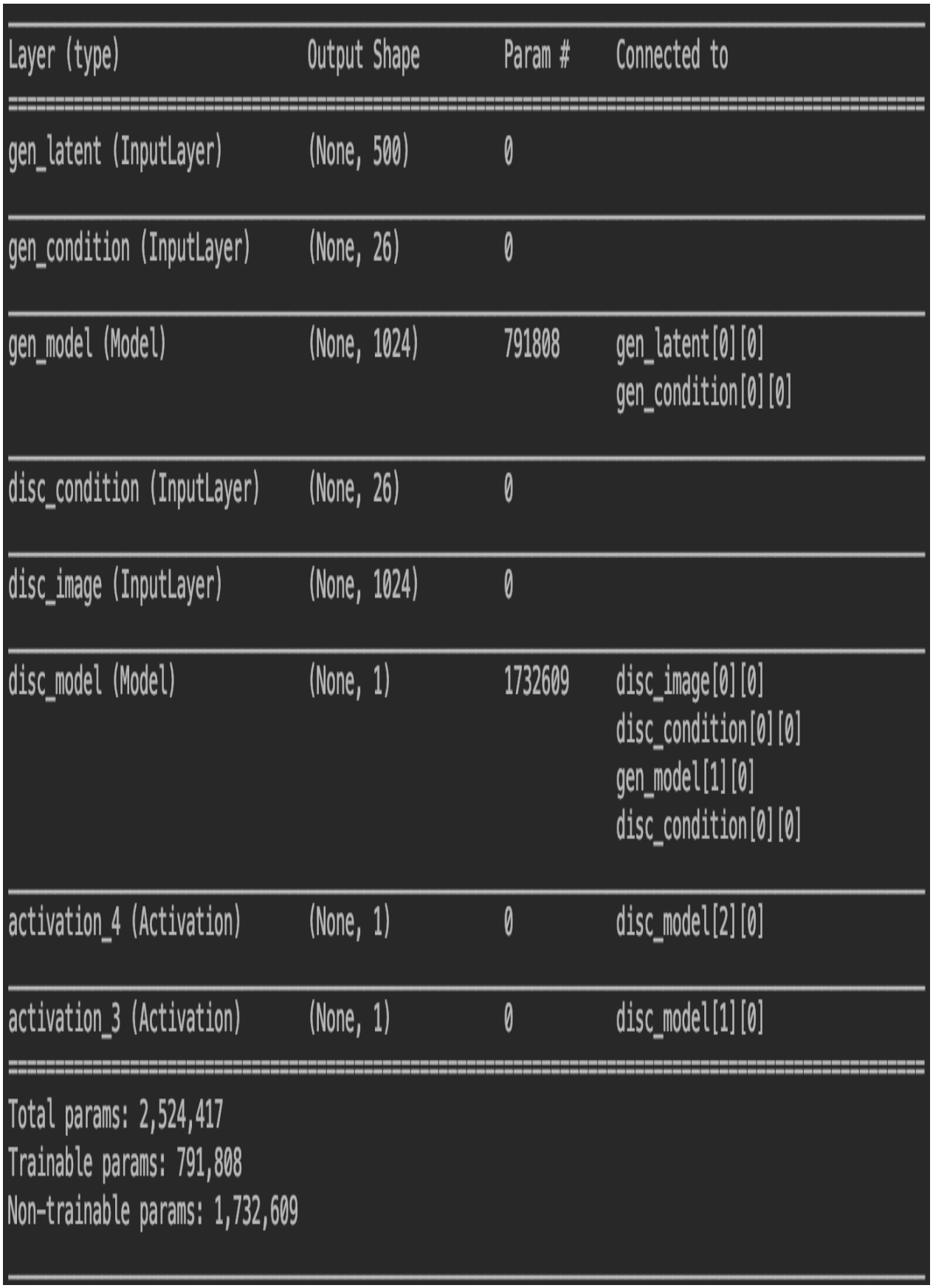
۴ توضیح معماری Discriminator

برای شبکه Discriminator موجود در معماری‌های CGAN و GAN از دو معماری مختلف استفاده شده است. یکی شبکه‌ای کانولوشنی که از لایه Image (به همراه لایه Condition در شبکه CGAN) و دیگری شبکه عصبی Fully Connected خالص که هر دو آزمایش شده‌اند و نتایج حاصل از آن‌ها موجود می‌باشد. برای شبکه CGAN،لایه Condition در حالت FullyConnected به لایه Image چسبانده (Concatenate) شده است. در واقع در این شبکه لایه Image ابتدا Flat شده است و یک لایه حاوی ۱۰۲۴ نورون را تشکیل داده است.شکل مربوط به معماری شبکه Discriminator در زیر آمده است.

Layer (type)	Output Shape	Param #	Connected to
disc_image (InputLayer)	(None, 1024)	0	
disc_condition (InputLayer)	(None, 26)	0	
concatenate_2 (Concatenate)	(None, 1050)	0	disc_image[0][0] disc_condition[0][0]
dense_4 (Dense)	(None, 1024)	1076224	concatenate_2[0][0]
leaky_re_lu_3 (LeakyReLU)	(None, 1024)	0	dense_4[0][0]
dense_5 (Dense)	(None, 512)	524800	leaky_re_lu_3[0][0]
leaky_re_lu_4 (LeakyReLU)	(None, 512)	0	dense_5[0][0]
dense_6 (Dense)	(None, 256)	131328	leaky_re_lu_4[0][0]
leaky_re_lu_5 (LeakyReLU)	(None, 256)	0	dense_6[0][0]
dense_7 (Dense)	(None, 1)	257	leaky_re_lu_5[0][0]
activation_2 (Activation)	(None, 1)	0	dense_7[0][0]

Total params: 3,465,218
Trainable params: 1,732,609
Non-trainable params: 1,732,609

شکل ۳: معماری شبکه Discriminator



Layer (type)	Output Shape	Param #	Connected to
gen_latent (InputLayer)	(None, 500)	0	
gen_condition (InputLayer)	(None, 26)	0	
gen_model (Model)	(None, 1024)	791808	gen_latent[0][0] gen_condition[0][0]
disc_condition (InputLayer)	(None, 26)	0	
disc_image (InputLayer)	(None, 1024)	0	
disc_model (Model)	(None, 1)	1732609	disc_image[0][0] disc_condition[0][0] gen_model[1][0] disc_condition[0][0]
activation_4 (Activation)	(None, 1)	0	disc_model[2][0]
activation_3 (Activation)	(None, 1)	0	disc_model[1][0]

Total params: 2,524,417
Trainable params: 791,808
Non-trainable params: 1,732,609

شکل ۴: معماری کلی شبکه CGAN در زمان آموزش شبکه Generator

۵ توضیح نحوه آموزش شبکه‌های مولد

برای آموزش شبکه‌های مولد، از Trickهای مختلفی استفاده شده است. این تریک‌ها در زیر آورده شده‌اند.

- استفاده از الگوریتم بهینه‌سازی Adam برای شبکه Generator و الگوریتم بهینه‌سازی SGD برای شبکه Discriminator
- استفاده از تکنیک Label Smoothing برای آموزش هر دو شبکه Generator و Discriminator
- استفاده از تابع فعال‌ساز LeakyReLU با پارامتر ۰.۲ بعد از هر لایه در معماری شبکه‌های Generator و Discriminator
- آموزش جدای داده‌های realو fake به شبکه Discriminator به جای آموزش ترکیبی از آنها (concatenationای از آنها)
- نرمالیزه کردن دیتای ورودی برای آموزش شبکه‌های مولد(نگاشت بازه هر پیکسل ورودی به ۱ و ۱-)
- آموزش داده‌های واقعی با برچسب صفر و داده‌های غیر واقعی با برچسب ۱ برای شبکه Generator

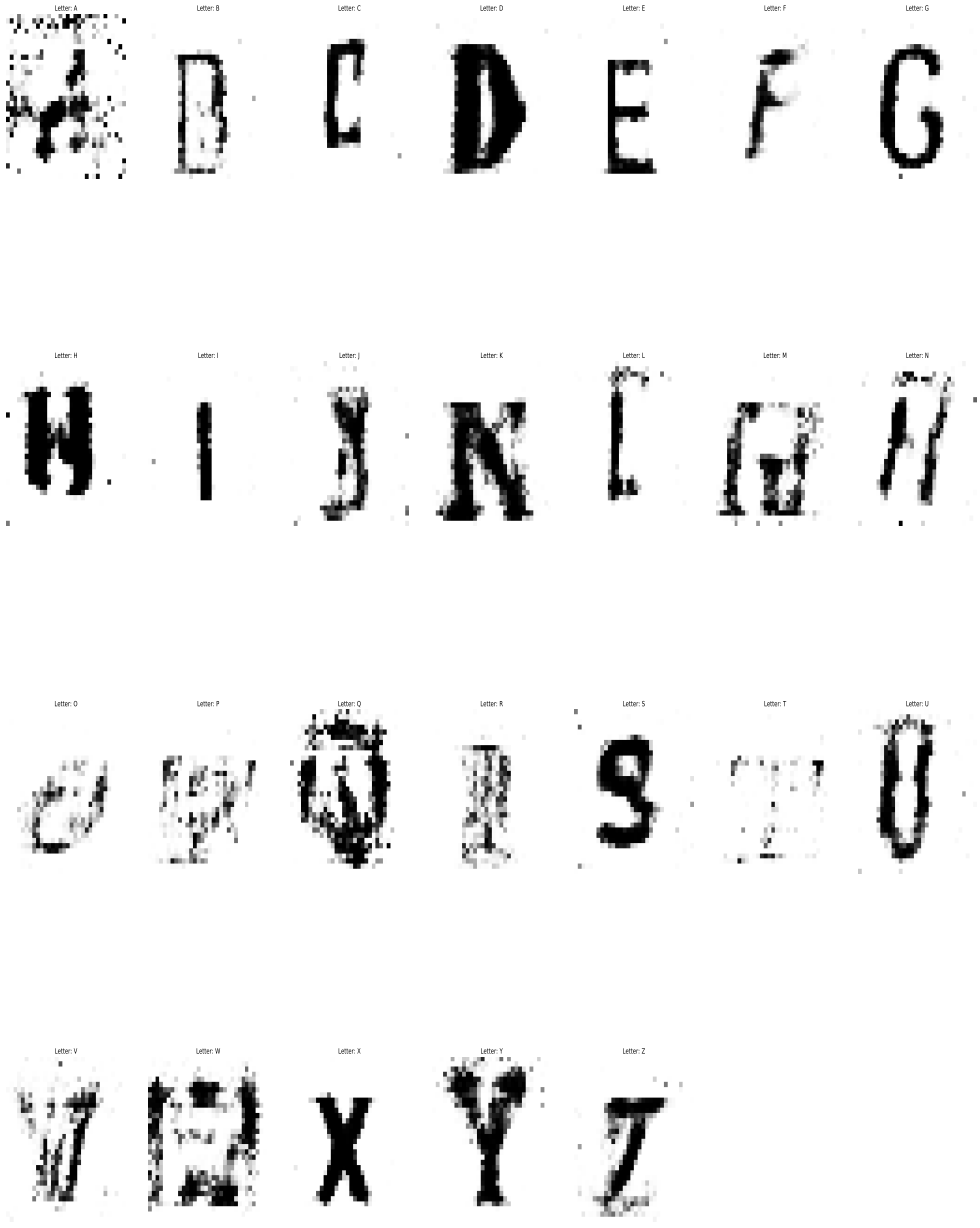
۶ نتایج بدست آمده در قسمت اول

برای شبکه اول، از تمام داده‌های Training برای آموزش استفاده شده است و از داده X_test داده شده برای داده Validation و ارزیابی شبکه استفاده شده است. نمودار یادگیری برای داده آموزش و تست برای این شبکه در زیر آمده است.

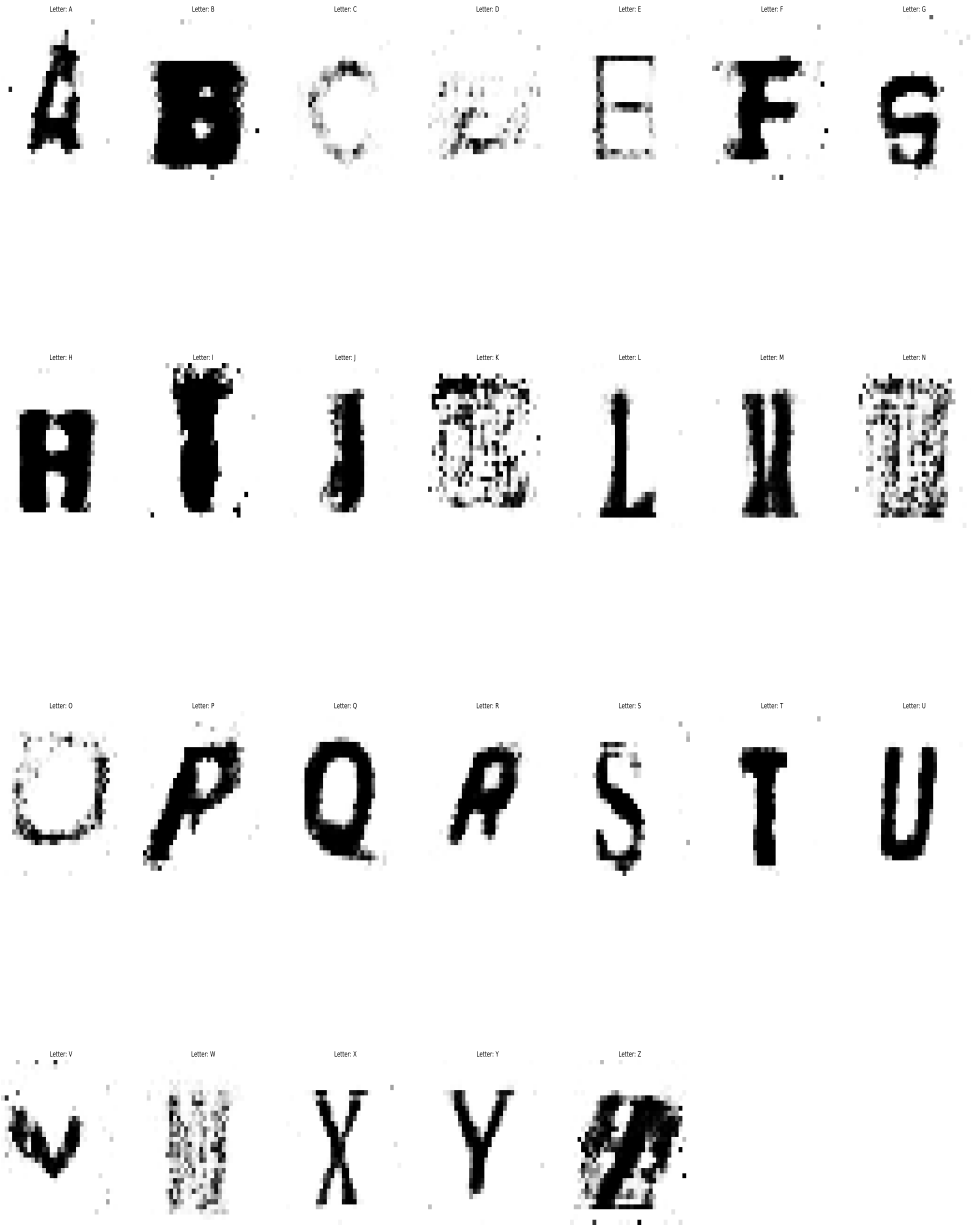
۷ نتایج بدست آمده در قسمت دوم

برای این قسمت نتایج عظیمی فراهم شده است! اکثر این نتایج در پوشه results ضمیمه شده‌اند. همچنین عکس‌های مربوط به تولید جمله مورد نظر در صورت پروژه و کاراکترهای مختلف در طول آموزش شبکه CGAN در زیر آمده است.

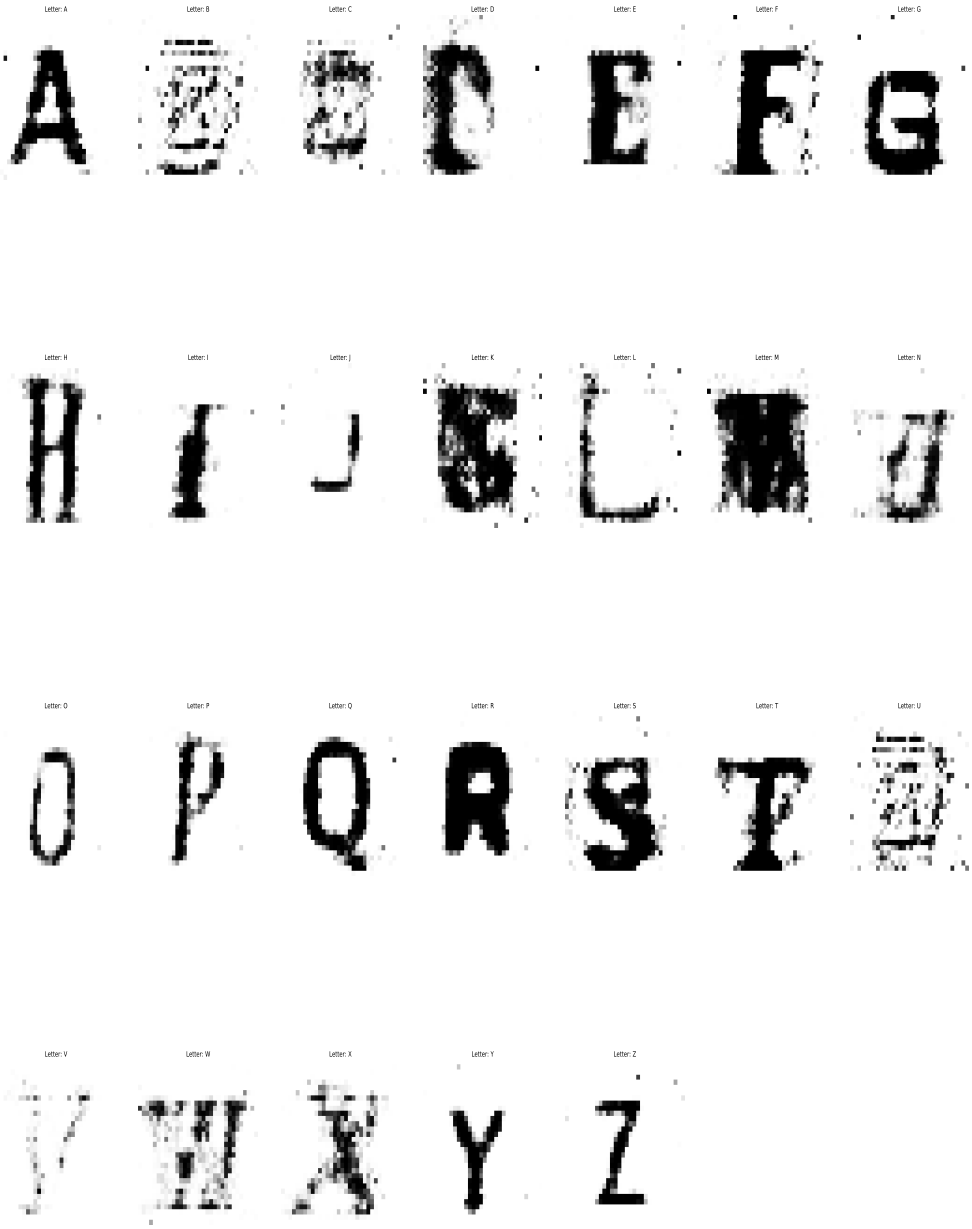
۱.۷ حروف الفبا



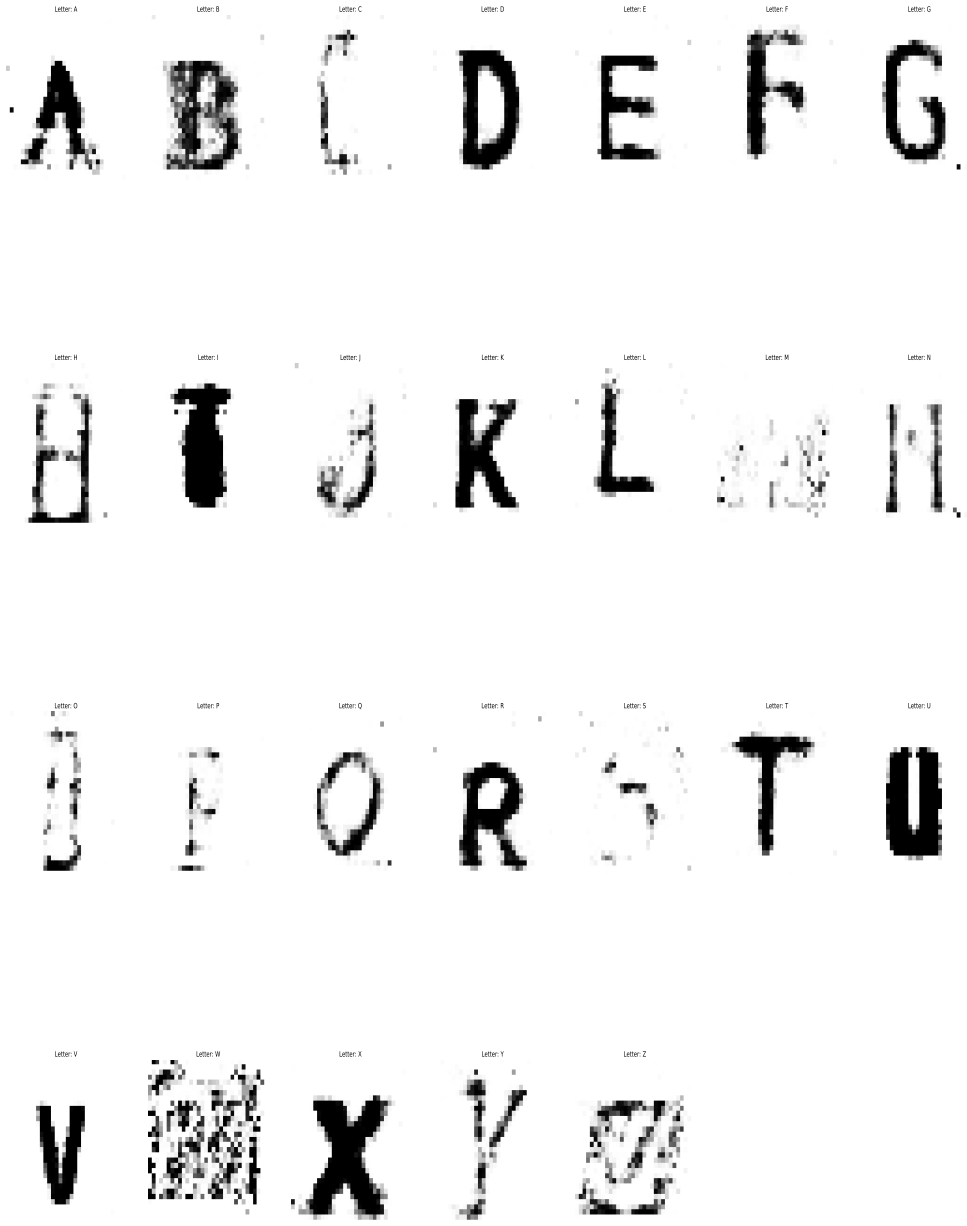
شکل ۵: عکس‌های الفبای یادگرفته شده در ایپاک ۱۷۵م در شبکه عصبی



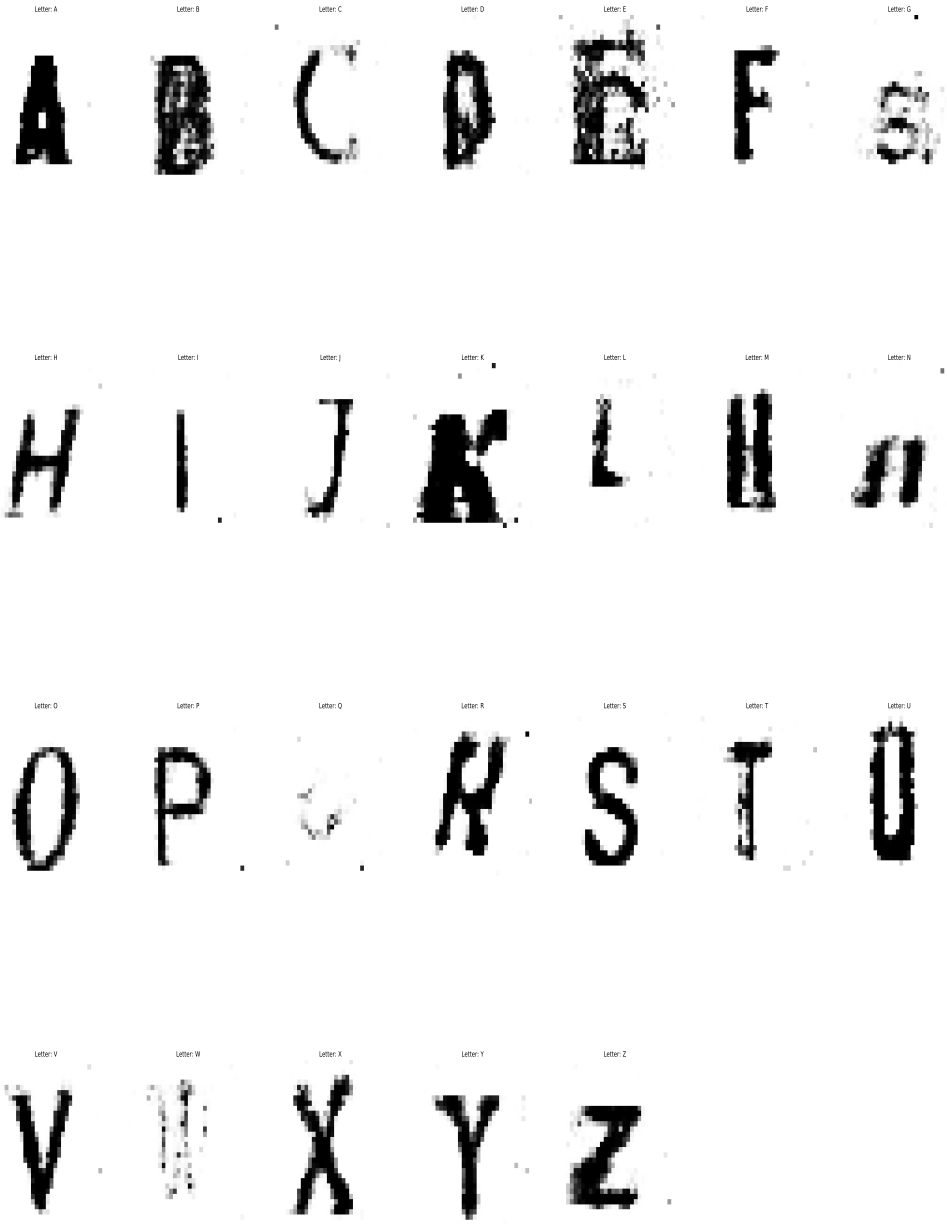
شکل ۶: عکس‌های الفبای یادگرفته شده در ایپاک ۸۰م در شبکه عصبی



شکل ۷: عکس‌های الفبای یادگرفته شده در ایپاک ۸۵م در شبکه عصبی



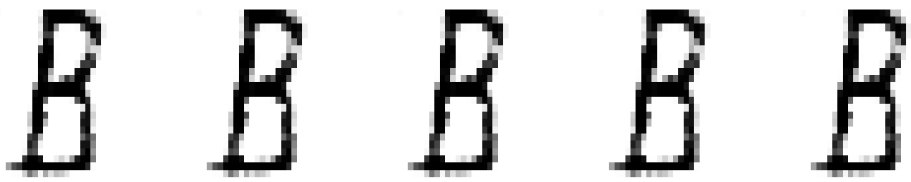
شکل ۸: عکس‌های الفبای یادگرفته شده در ایپاک ۹۰م در شبکه عصبی

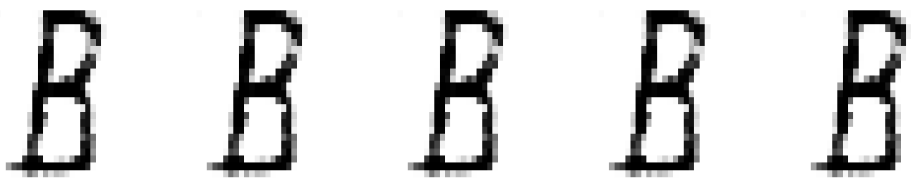


شکل ۹: عکس‌های الفبای یادگرفته شده در ایپاک ۹۵م در شبکه عصبی

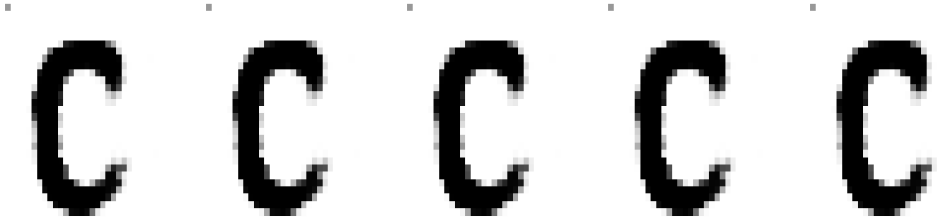
۲.۷ حرف A

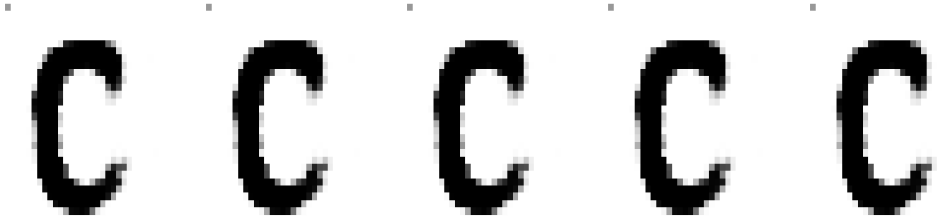
۳.۷ حرف B

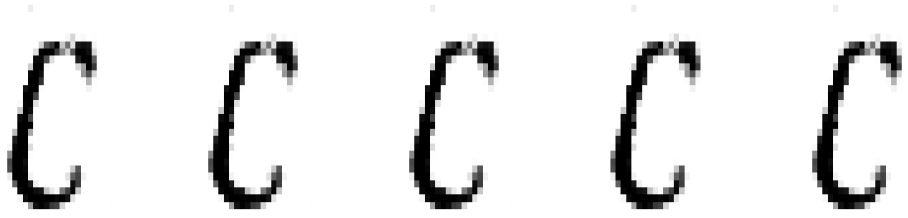


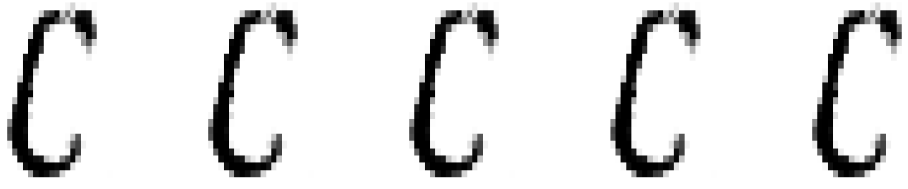


۴.۷ حرف C

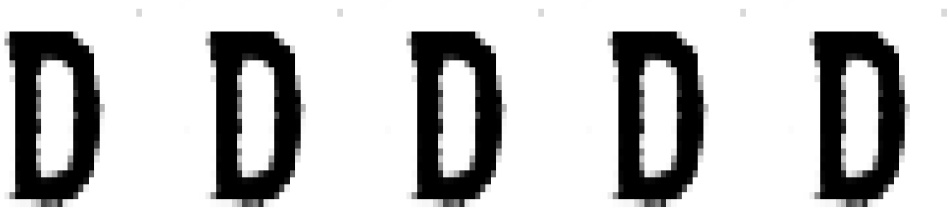


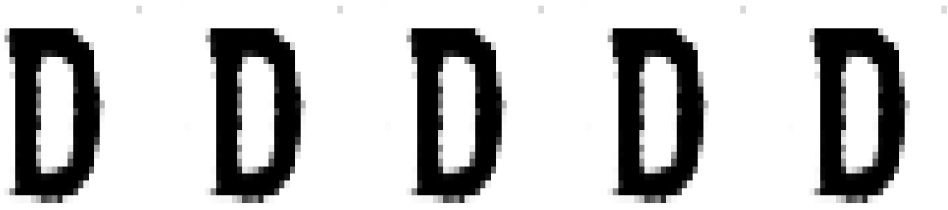


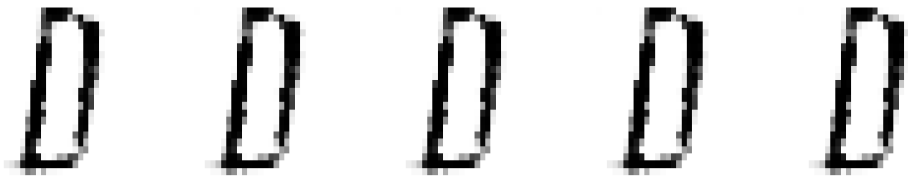


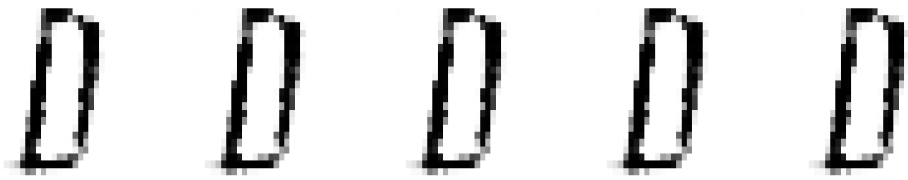


۵.۷ حرف D

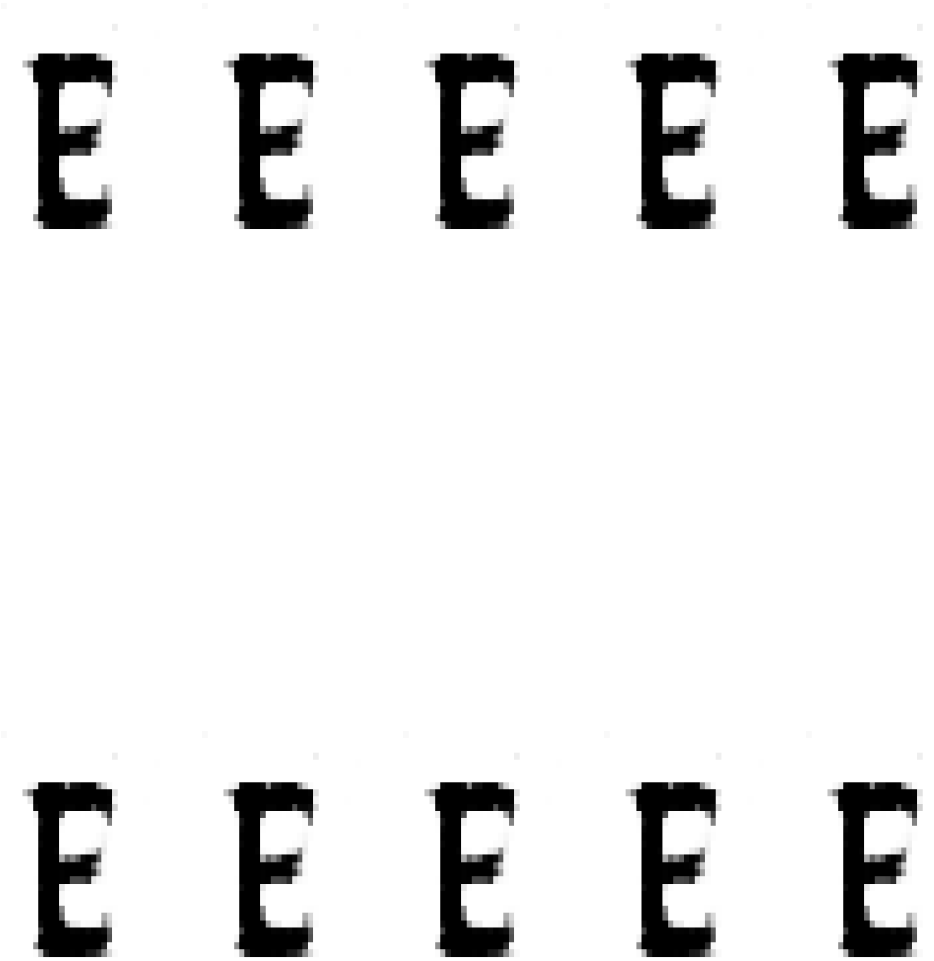


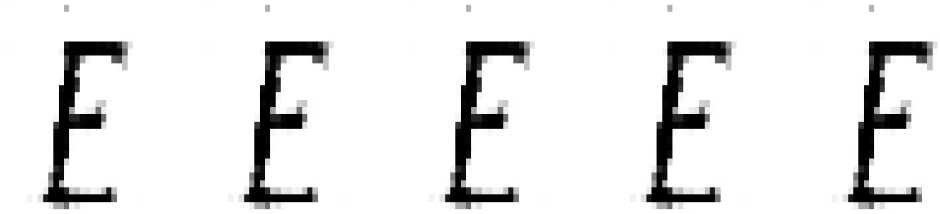


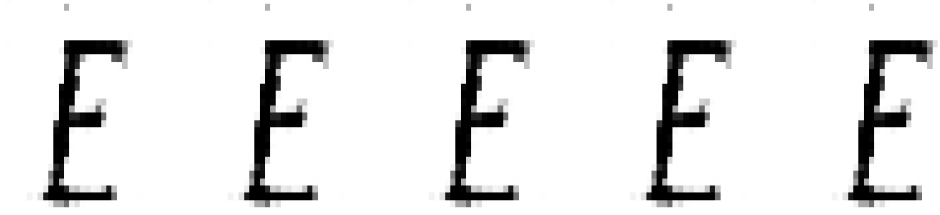




۶.۷ حرف E



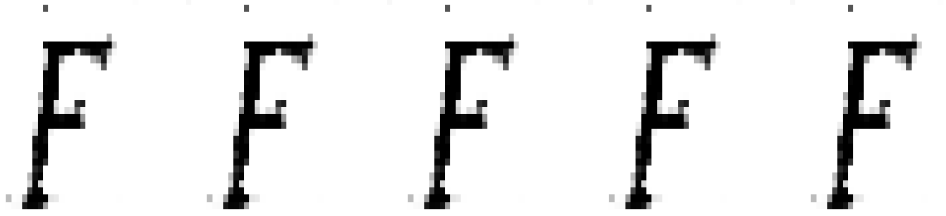




۷.۷ حرف F

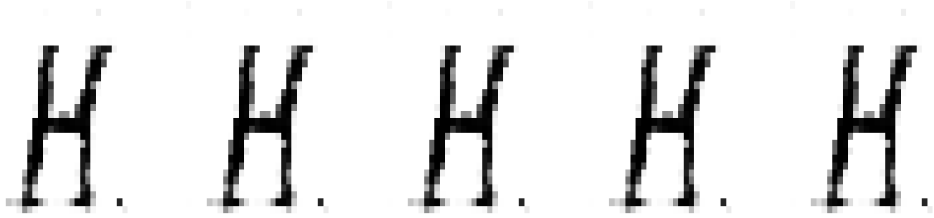
F F F F F

F F F F F

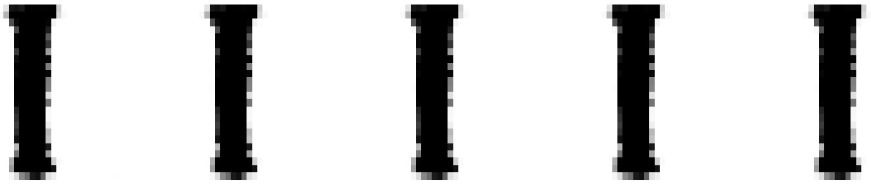


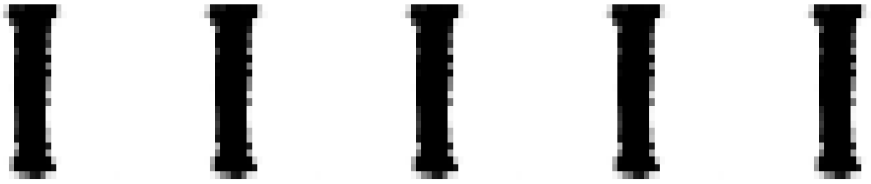
G G G G G

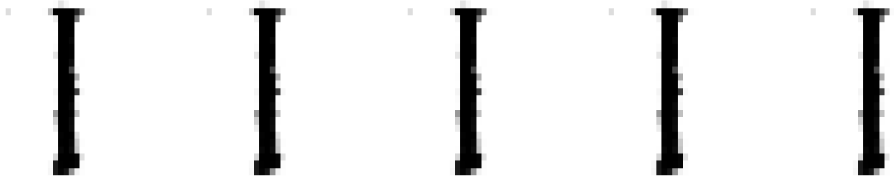
G G G G G



۱۰.۷ حرف I



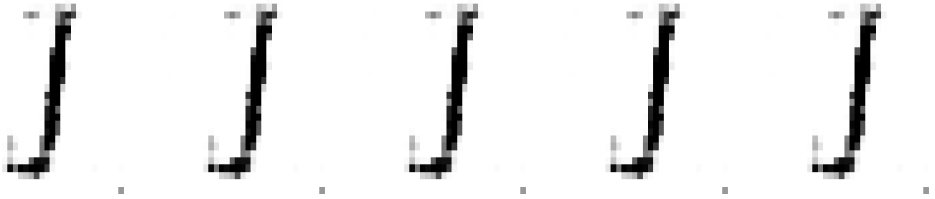


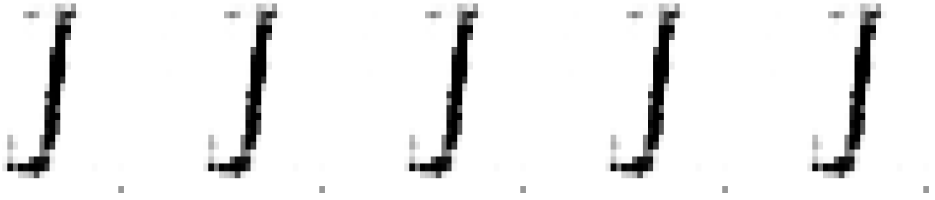


۱۱.۷ حرف J

J J J J J

J J J J J





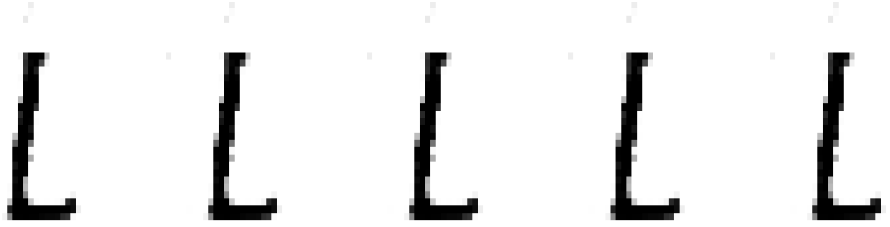
K K K K K

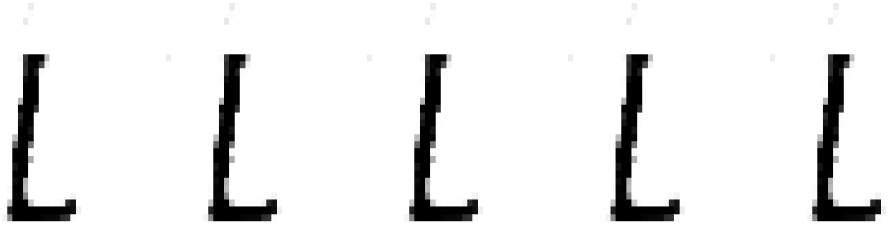
K K K K K

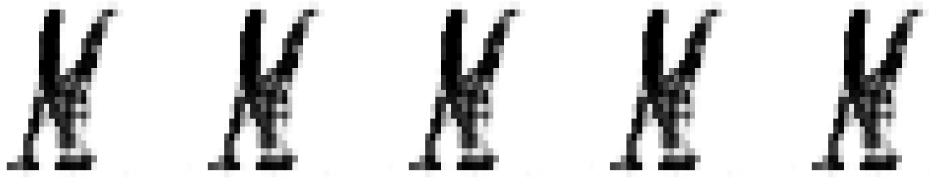
۱۳.۷ حرف L

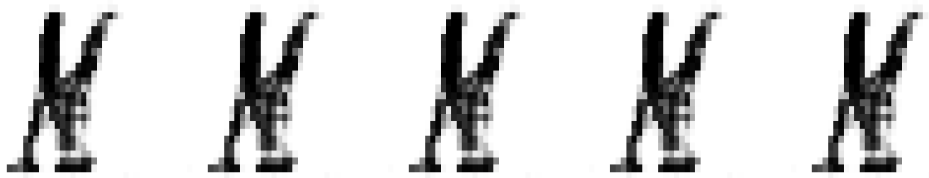
L L L L L

L L L L L

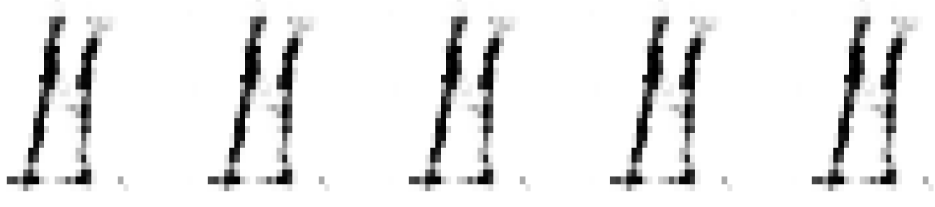


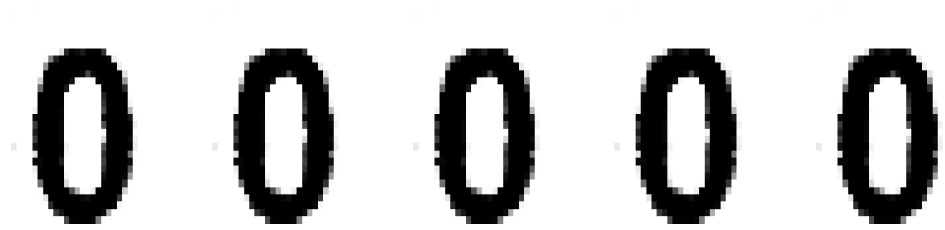


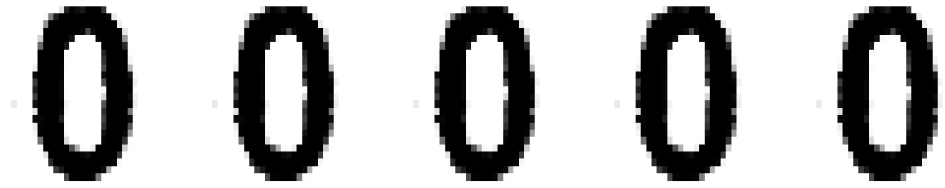


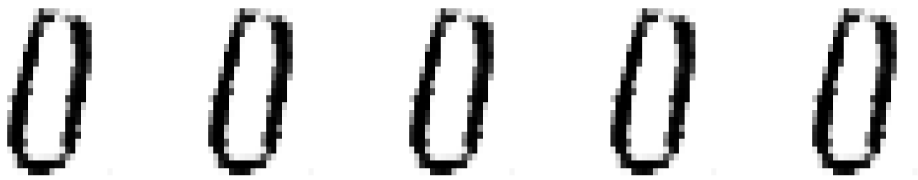


۱۵.۷ حرف N

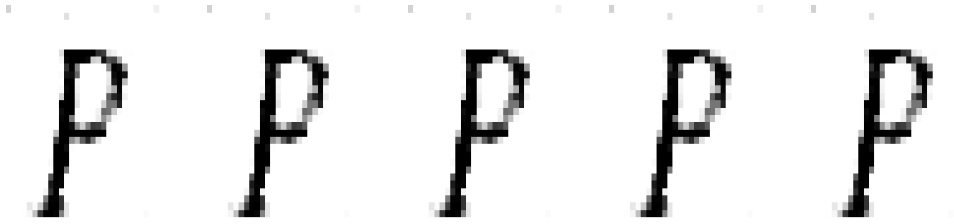


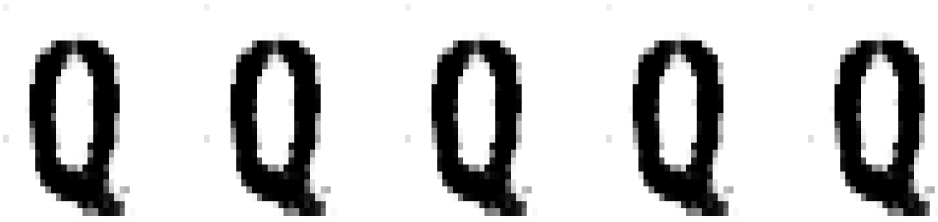


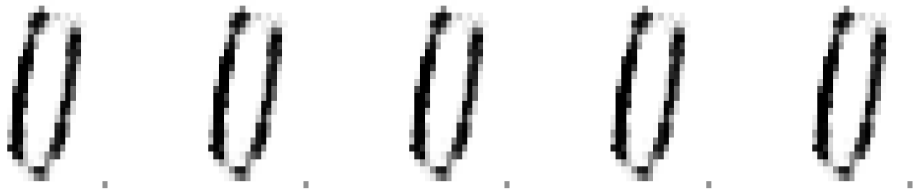


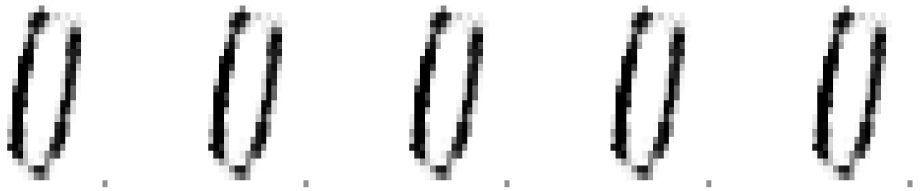


۱۷.۷ حرف P









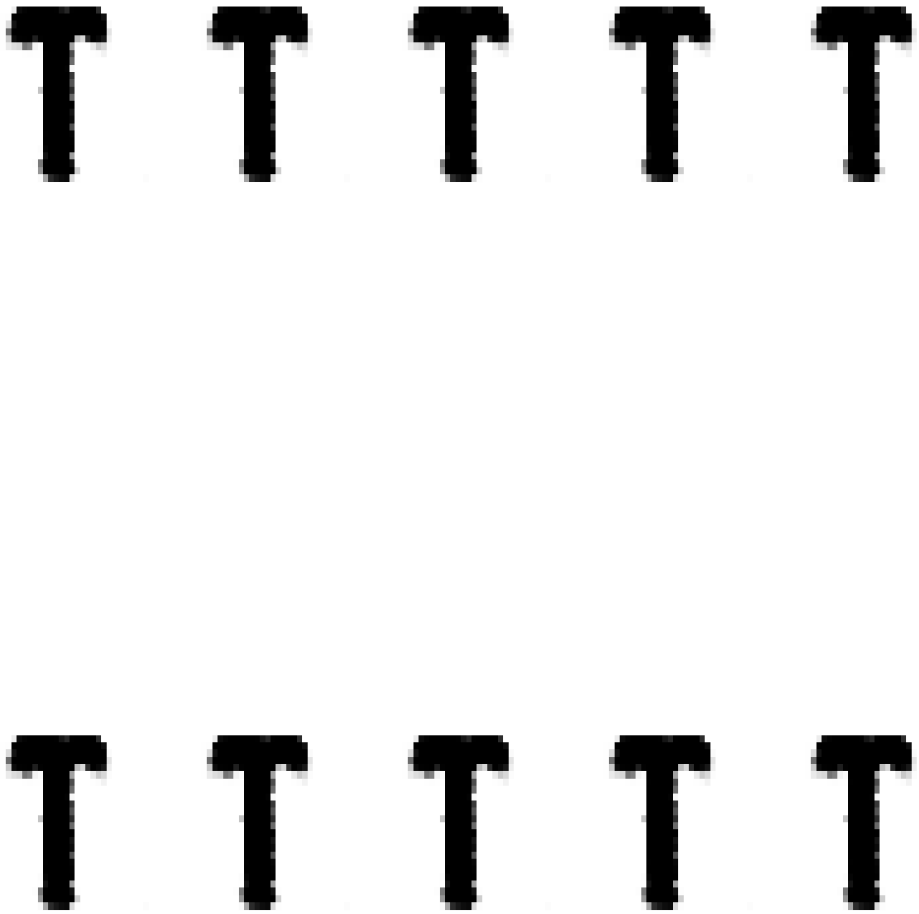
R R R R R

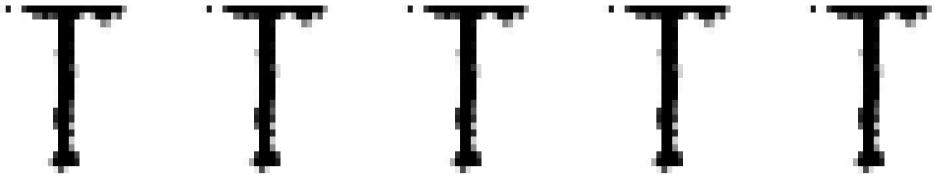
R R R R R

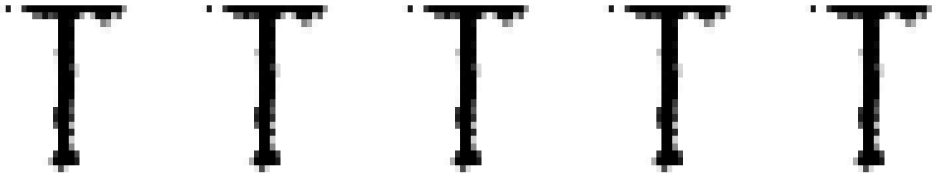
S S S S S

S S S S S

۲۱.۷ حرف T



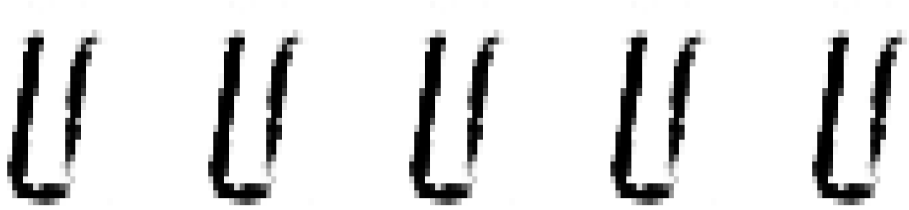


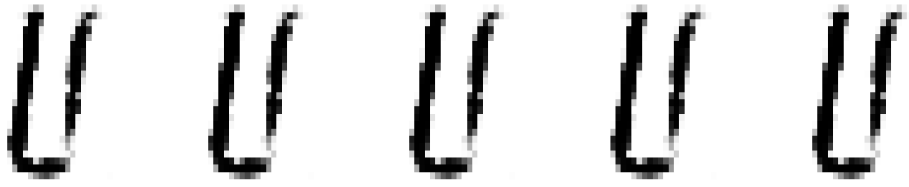


۲۲.۷ حرف U

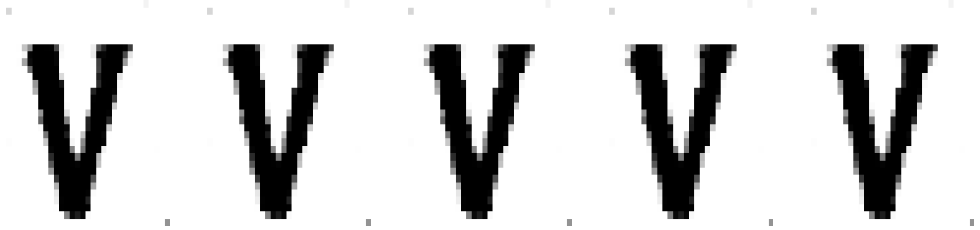
U U U U U

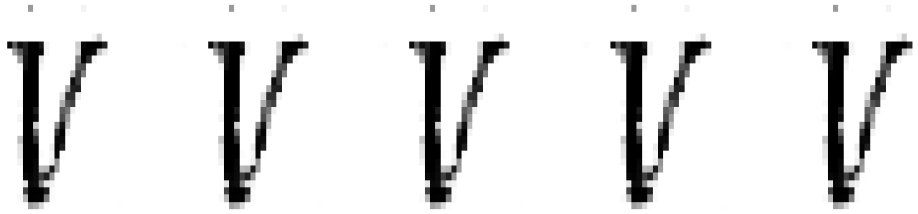
U U U U U

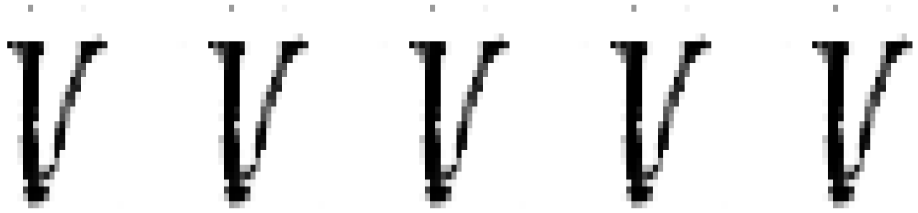




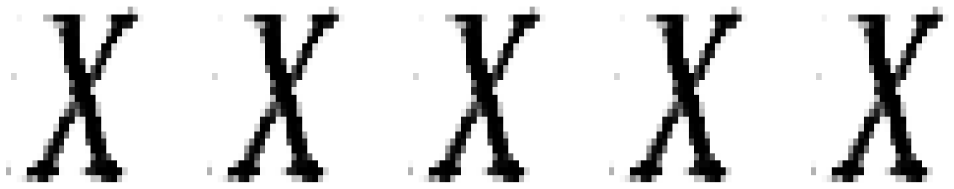
۲۳.۷ حرف V

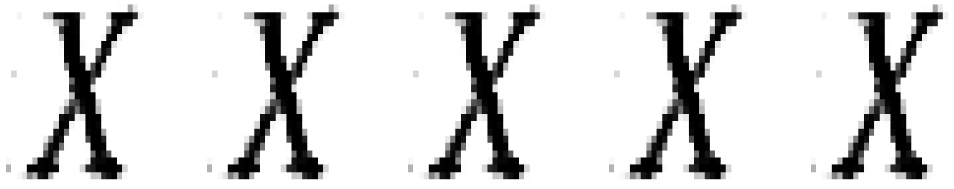






۲۵.۷ حرف X





Y Y Y Y Y

Y Y Y Y Y

z z z z z

z z z z z