

تمرین 4 یادگیری عمیق

توضیحات سوال 4

امیرعباس افضلی 400100662

تفاوت و مزیت های VQ-VAE و VAE :

VQ-VAE:

1. VQ-VAE یک مدل تولید متغیر است که از ساختار کوانتایزرها و کتوری برای فشرده سازی و مبدل خود استفاده می کند.
2. یکی از مزایای VQ-VAE این است که می تواند اطلاعات را در فضای کم بعدی ترکیب کند و باعث ایجاد تصاویر با ویژگی های نوین (همانند بخش آخر سوال) شود.

و همچنین از دیگر مزایای VQ-VAE ، توانایی کاهش ابعاد داده ها است که به کاهش حجم حافظه مورد نیاز برای ذخیره و پردازش داده ها کمک می کند. این الگوریتم همچنین می تواند ویژگی های مهم داده را استخراج کرده و از اطلاعات معنایی آن ها حفاظت کند. از دیگر مزایای این روش می توان به امکان بازسازی داده های ورودی، کاهش افت اطلاعات در فرآیند کوانتیزه کردن و افزایش سرعت آموزش اشاره کرد.

VAE:

1. VAE نیز یک نوع مدل تولید متغیر است که از شبکه های عصبی برای یادگیری نمایشی از داده ها استفاده می کند که از لحاظ احتمالاتی تجسم پذیر باشد.
2. یکی از مزایای VAE این است که از تئوری احتمالات برای آموزش مدل استفاده می کند و اطمینان حاصل می کند که نمایش های یادگرفته شده معقول و پذیرفتنی باشند.
3. VAE می تواند نمایش هایی از داده ها ایجاد کند که قابلیت تولید داده های جدید مشابه به داده های اصلی را داشته باشند.

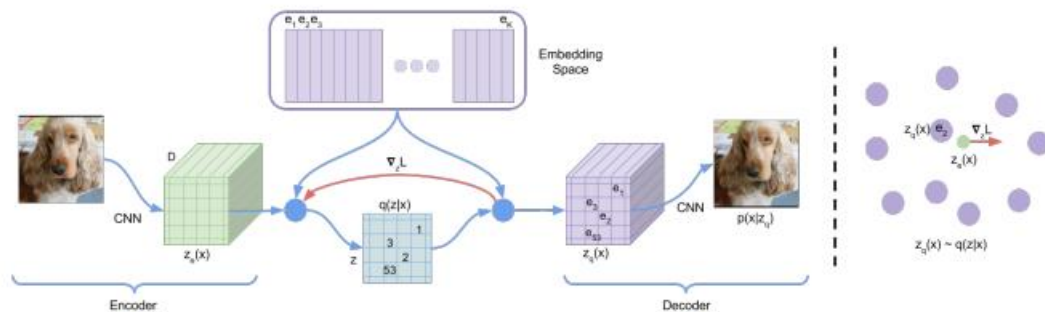


Figure 1: Left: A figure describing the VQ-VAE. Right: Visualisation of the embedding space. The output of the encoder $z(x)$ is mapped to the nearest point e_2 . The gradient $\nabla_z L$ (in red) will push the encoder to change its output, which could alter the configuration in the next forward pass.

در پیاده سازی شبکه نیز تا حد زیادی از شبیه مدل گفته شده در مقاله استفاده شده است :

4.1 Comparison with continuous variables

As a first experiment we compare VQ-VAE with normal VAEs (with continuous variables), as well as VIMCO [28] with independent Gaussian or categorical priors. We train these models using the same standard VAE architecture on CIFAR10, while varying the latent capacity (number of continuous or discrete latent variables, as well as the dimensionality of the discrete space K). The encoder consists of 2 strided convolutional layers with stride 2 and window size 4×4 , followed by two residual 3×3 blocks (implemented as ReLU, 3×3 conv, ReLU, 1×1 conv), all having 256 hidden units. The decoder similarly has two residual 3×3 blocks, followed by two transposed convolutions with stride 2 and window size 4×4 . We use the ADAM optimiser [21] with learning rate $2e-4$ and evaluate the performance after 250,000 steps with batch-size 128. For VIMCO we use 50 samples in the

Effect of number of embeddings

همانطور که در نوت بوک هم آمده است مشاهده می شود که با افزایش تعداد بردار ها مقدار training loss نهایی که مدل به آن کانورج می کند کاهش می یابد . در خروجی هم مشاهده می شود که با افزایش تعداد بردار ها ، کیفیت رنگ آمیزی کمی بهبود می یابد .

Effect of dimension of embeddings

در این بخش مشاهده می شود که با افزایش بعد بردار ها مقدار training loss نهایی که مدل به آن کانورج می کند بدتر (بیشتر) می شود . در خروجی هم مشاهده می شود که با افزایش بعد بردار ها ، کیفیت رنگ آمیزی بدتر می شود و خروجی ها کمی blur و کمرنگ خواهند شد .