

Share



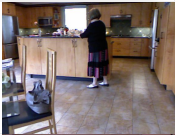

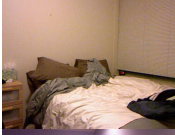
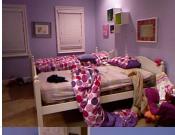
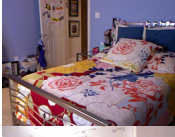

# VQA as Encoder-Decoder with ViT and GPT-2 as backbones

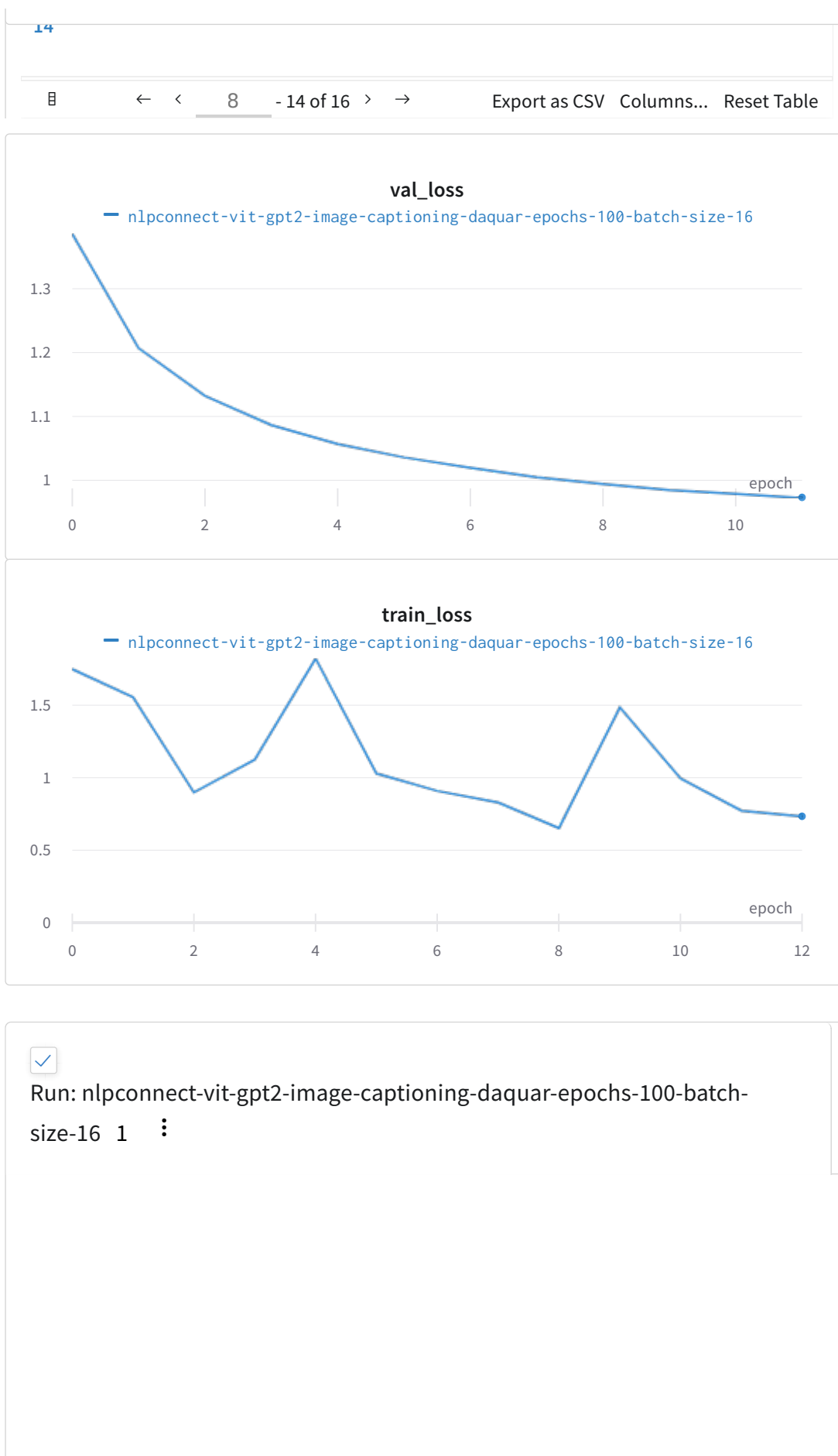
The backbone has been pre-trained on the image captioning task. It needs more fine-tuning.

Vladyslav Fedoriuk

## ▼ Section 1

`runs.summary["sample_predictions"]`

	Image	Question	Answer	Model prediction
8		what is the object found on the table	bag	answer the following question: what is the object found
9		what are the colours identifiable on the whiteboard	black	answer the following question: what are the colours
10		what is left side of the bed	night_stand	answer the following question: what is left side of the
11		what is the object close to the wall left of the window	picture	answer the following question: what is the object close
12		what is on the bed	pillow	answer the following question: what is on the bed
13		what is the object close to the wall behind the chair	whiteboard	answer the following question: what is the object close



Created with  on Weights & Biases.

<https://wandb.ai/vladfedoriuk/vqa/reports/VQA-as-Encoder-Decoder-with-ViT-and-GPT-2-as-backbones--Vmlldzo0NDM4OTQ3>