VQA



Gera Versfeld • 06.13.2017

The VQA-v2 Dataset

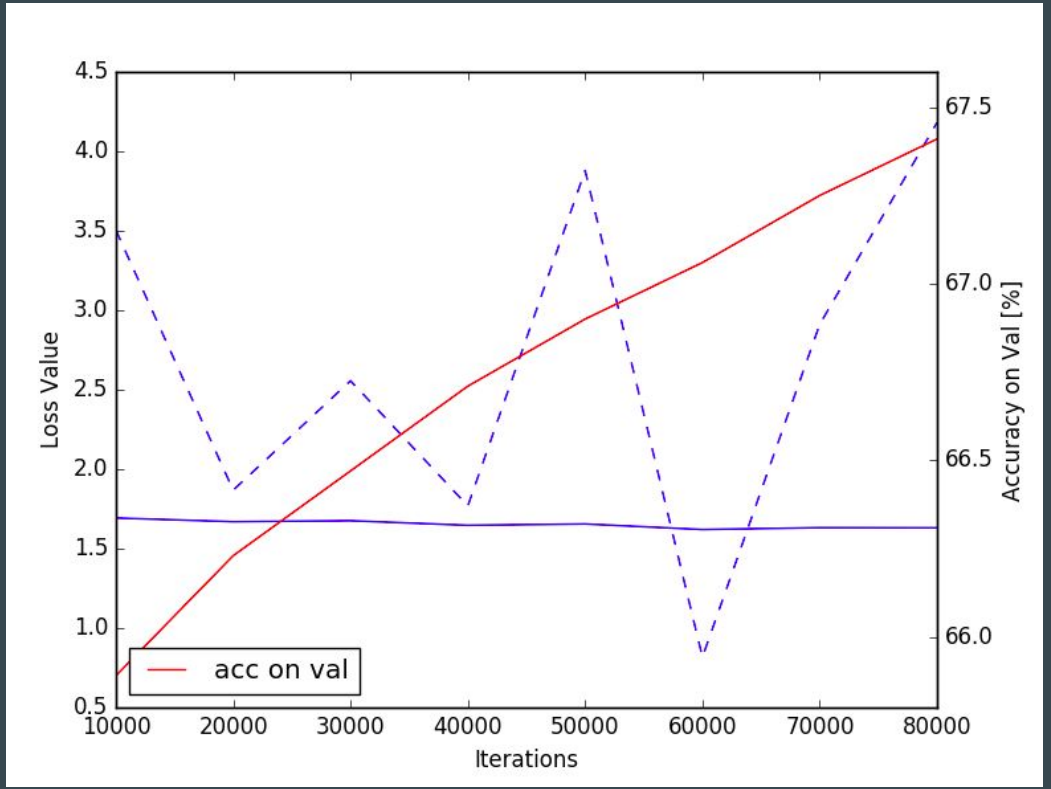
- Training: 4,437,570 questions, 443,757 answers, 82,783 images
- Validation: 2,143,540 answers, 214,354 questions, 40,504 images
- Test: 447,793 questions, 81,434 images

- MSCOCO images
- Balanced - answer for each question depends on image
- Top question types:
  - What - 41.2%
  - Is - 26.1%
  - How many - 10.3%
  - Are - 7.57%

http://visualqa.org/visualize/

Progress this week:

- Fine-tuning MCB baseline - highest accuracy on validation: 67.41%. Lowest test loss at 60,000 iterations
- Fine-tuning MCB baseline with residual loss - still training



Goals for next week:

- Test baseline on test-dev data set
- Find models better at answering ‘how many’
- Implement question attention scheme and answer regression
- Begin training
- Create ensemble model