Validation metrics over training

Validation loss

- Fine-tune (single concepts)
- Train from scratch
- Fine-tune head (summaries)

Top-10% retrieval

- Shuffle within batch
- Fine-tune (abstract)
- Fine-tune (summaries)

Random expectation