Original Data2Vec uses random crop in [0.2, 1.0] (or for the version used in BEiT 1(?)\*, which is apparently not used [0.08, 1.0], which is generally too strict!)

This is fine if goal is just to learn image representations, but not good for Image-Text pairs, like in COCO captions! -> Crop too harsh will discard/destroy some information in the image the fitting text/caption is about!

See src/datasets/coco\_crop\_problem.ipynb

\*should be the case, as BEiT 2/3 is multimodal! (maybe check that)