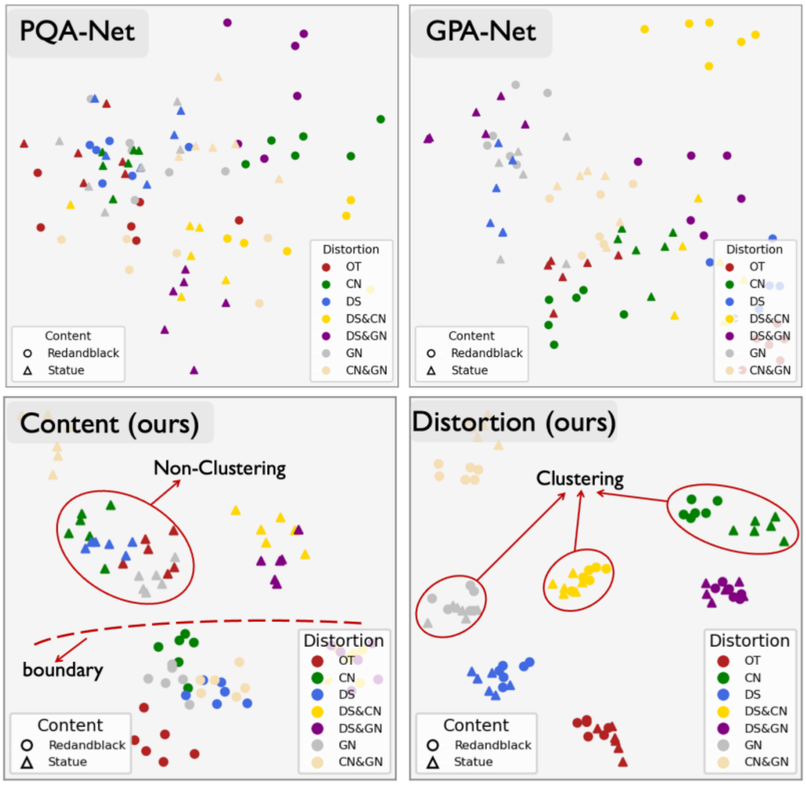

Figure A: **Visualization of t-SNE** embeddings of representations of PQA-Net, GPA-Net, content-aware and distortion-aware branches of our DisPA on the testing set of SJTU-PCQA. The scattered points are color and shape marked according to distortion type and content. The distortion-aware features are visualized in lower right, where clusering of distortion types exists clearly. The content-aware features present non-clustering for distortion types but a clear boundary to split the content.

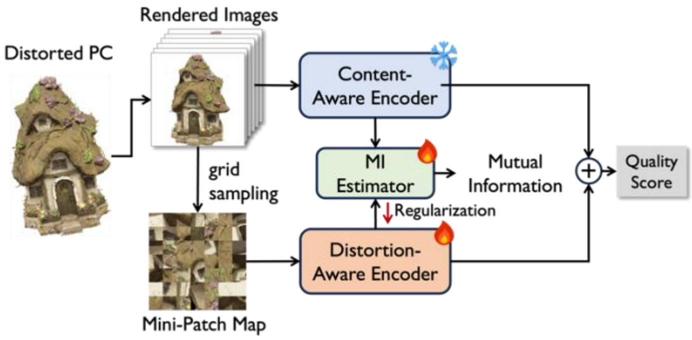

Figure B: Architecture of proposed DisPA. Our DisPA consists of two encoders $\mathcal{F}$ and $\mathcal{G}$ for learning content-aware and distortion-aware representations, and an MI estimator $\mathcal{M}$. Given a distorted point cloud, we first render it into multi-view images which are fed into $\mathcal{F}$ generate the content-aware representation. Next, the multi-view images are decomposed into a mini-patch map through grid mini-patch sampling. The mini-patch map is encoded by $\mathcal{G}$ obtain distortion-aware representation. After obtaining the representations, we use them to train the MI estimator $\mathcal{M}$..

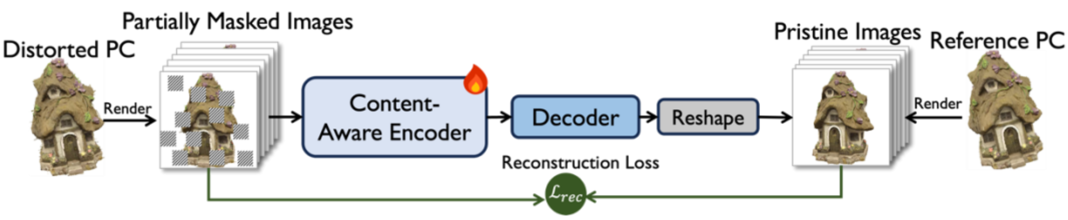

Figure C: Illustration of content-aware pretraining based on masked autoencoding. The pretraining maskes a part of patches and reconstruct them. The reconstruction loss is computed between reconstructed images and the rendered pristine images projected from reference point clouds.