* Meine Einschätzung: Ich glaube sie haben kein latent diffusion genutzt und deshalb brauchen sie so viele Ressourcen -> für nichts das Paper
* Ziel
  + Noise Diffusion haben Sota Performance. Brauchen jedoch viel Performance. Dies limitiert Einsatz für e.g. 3D Volumes
  + Sie möchten verschiedene Möglichkeiten zur Reduktion von Resources testen
  + Main contribution ist patch-based diffusion model PatchDDM
* Einführung
  + 3D Volumes waren soweit restricted zu 128x128x128
  + Jedoch nicht für höhere Qualität wie 256x256x256
* Anpassungen an 3D model
  + Es gab bereits 3D U-Net models, welche sie anpassten
  + Sie reduzierten self attention layers, weil dies weniger Ressourcen braucht
  + Sie nutzten additive skip connections
    - In originaler UNet Architektur waren skip connections umgesetzt als concatenations
    - Dies bedeutet dass Decoder mehr Ressourcen braucht als Encoder
    - Sie nutzten ein Averaging. Dies ist besser als Addition wie ResNet um Exploding Gradient zu verhindern
  + Für bessere Qualität haben sie Anzahl an Channels bei einem Faktor von 1.61 erhöht
* PatchDDM
  + Training erfolgt mit Patches
  + Inference aber trotzdem mit gesamtem 3D Volume möglich
  + Sie condition Netzwerk mit Position vom sampled Patch
* Ablation study
  + Vergleich mit zwei Baselines ohne patch-based training
    - Mit Full Resolution
      * Dazu brauchten sie zwei GPUs
    - Mit half Resolution
      * Sie downsampled images von 256^3 zu 128^3, gaben es in Model und upsampled Output wieder
      * -> Reduktion um Faktor 8 für memory/computation time
  + Sie versuchten auch ohne coordinate encoding, jedoch hat dies nicht converged
* Preprocessing
  + Sie padden Scans von 240x240x155 zu 256x256x256
  + Background voxels wurden zu 0 gesetzt
  + Range zwischem 1st und 99th Percentile wurde normalisiert zu [0,1]
* Training
  + Nvidia A100 GPU mit 40 GB Memory
  + Training für 420h (2.5 Wochen)
  + Memory consumption
    - FullRes brauchte 80 GB
    - HalfRes und PatchDDM brauchten 12GB
  + Für Inference brauchte PatchDDM memory und speed ähnlich zu FullRes
* Discussion
  + Qualität ist nicht so gut wie state of the Art Unet (nnUnet)
    - <https://www.nature.com/articles/s41592-020-01008-z> -> sota unet
  + Ein Bild, das Text, Schrift, Screenshot, Zahl enthält.

    Automatisch generierte Beschreibung
  + Memory Consumption und Training time konnte sicher reduziert werden
  + Mir ist aber nicht klar, ob auch overall Training Time schneller war