Evaluation metrics

* PSNR
  + Peak signal to noise ratio
  + Vergleicht reine pixel values und nicht Dinge, welche das menschliche Auge in Betracht ziehen würde
  + Ein Bild, das Text, Schrift, weiß, Typografie enthält.

    Automatisch generierte Beschreibung
  + MSE ist mean squarred error zwischen zwei Bilder
  + MAX\_f ist maximum possible pixel value, welches im «guten» Bild existiert (Bei 8bits per sample wären es 255)
  + Grosse Nummern bedeuten, dass Bilder ähnlicher sind.
* SSIM
  + Verbesserung von PSNR
  + Ein Bild, das Text, Schrift, Handschrift, weiß enthält.

    Automatisch generierte Beschreibung
    - c1, c2 sind für nummerische Stabilität
  + Diese Formel steht für Vergleich von Helligkeit l, Kontrast c und Struktur s
    - Ein Bild, das Text, Schrift, Handschrift, Quittung enthält.

      Automatisch generierte Beschreibung
    - 
    - Alpha, Beta und Gamma sind Weights
  + Gibt mir ein Wert zwischen -1 und 1
    - 1 ist perfekte similarity
    - 0 ist keine similarity
    - -1 ist perfekte anti-correlation
* FID
  + Funktioniert anscheinend nicht gut für stable diffusion models
  + Man gibt Bilder durch Inception-v3 model
  + Man berechnet mean und covariance matrix von features in jedem Bild
  + Man berechnet die distanz zwischen mean und covariance matrix zwischen den Bildern
  + Kleinere Nummern bedeuten dass Bilder ähnlicher sind.
* LPIPS
  + Gut beschrieben in Literatur 05 Machine Learning
  + Verbesserung zu SSIM und PSNR
  + Als metric nehmen sie auch deep features von DL Netzwerken genommen werden (e.g. VGG) und dann die l2 distance von diesen Features als perceptual loss. Sie machen es aber über alle intermediate features im Netzwerk
    - Man hat zwei Bilder und lässt diese durch ein Netzwerk
    - Der resultierende Feature stack von beiden Bildern wird entlang der channel Richtung normalisiert und dann subtrahiert
    - Anschliessend l2 norm nehmen und dann über spatial dimension normalisieren (l2 pro neuron berechnen und averagen)
    - Dann am Schluss nochmals über alle Layer averagen
    - Ein Bild, das Diagramm, Reihe, Screenshot, Text enthält.

      Automatisch generierte Beschreibung
  + Sie trainieren noch eigene Netzwerke, welche die besten Features für LPIPS ergeben

Papers und genutze Metrics

* Stable Diffusion / Lama
  + Gemäss Protokoll von Lama
  + FID
  + LPIPS
  + Human evaluation
  + Ich glaube Metriken basierend auf Patches berechnet
* Inpaint Challenge
  + PSNR
  + SSIM
  + MSE

TODO

* Vergleiche ich nur den inpainted Bereich oder das gesamte Bild?