Paper summary

AI VISION Lab

1. 공부한 논문의 제목, 게재된 학회 혹은 저널 등 논문 기본 정보를 적으세요.
   - **A. 이름: SR3: Image Super-Resolution via Iterative Refinement**
   - **B. 저널: IEEE TPAMI**
   - **C. 도메인: Super Resolution**
   - **D. 출판연도: 2021**
   - **E. 저자: Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, Mohammad Norouzi (Google Research, Brain Team)**

2. 논문에서 제안한 알고리즘 및 프레임워크에 대해 본인이 이해한대로 다이어그램을 그려보세요. **논문 Figure를 그대로 따라 그리면 안됩니다.**
   A. 일반적인 구조는 Diffusion(DDPM)과 동일함 (UNet사용, 충분한 T Size, step 당 Noise 크기는 최소화)

   B. 다만 차이점은 다음과 같음. 우선, Low Resolution Image x에 대해서 Resample을 진행함. (목표는: High Resolution Image와 동일한 크기의 이미지로 맞춰주는 것. Bicubic Interpolation -> 해당 Interpolation 기법의 경우 다른 SR에서도 자주 사용됨.)

   C. 이때 Resampled LR 과 y_t(이때 y_t의 경우- Reversed process -> Gaussian Noise인 y_T에서 Iterative하게 진행 -> Iterative Refinement)을 Concat (Concat말고 다른 방식도 시도해봤으나 Concat대비 성능이 나오지 않았다고 함– FiLM)

   D. 이후 학습 진행 (Diffusion과 동일, 이때 모델의 목표는 $p(x)$가 아닌 $p(y|x)$ 라는 점을 반영하여, x를 condition으로 주입하는 구조임. x: source | y = target) 마치 x0를 condition으로 주는 DDIM의 구조와 상당히 유사함. 다만 Forward 과정을 x0 conditioning하게 진행하지 않았다는 점 -> 여전히 Markovian하다고 볼 수 있음.)

   E. 추가적으로 Cascaded에 대해서도 다뤘는데, 즉각 High resolution으로 진행하는 것보다, 64 -> 256 -> 1024(Model 두 번)로 진행하는게 비용, 성능 측

면에서 장점이 있음을 보였음.

3. 본인이 생각하는 이 논문의 장점이 무엇이라고 생각하나요? **논문 Contribution bullet을 그대로 따라 적으면 안됩니다.**
   A. Condition, Uncondition, Face, Natural 다양한 조건에서 실험을 진행했고 성능을 입증했다는 점이 가장 큰 장점이 아닐까 생각됨.

4. 이 논문을 읽으면서 느낀 점, 혹은 배운 점이 있으면 적어보세요.
   A. 실제 사람의 판단을 기반으로 평가하는 점: (2AFC)
   B. SR의 목표 Distribution. 그리고 Regression(SR model)과의 차이점

5. 이 논문의 한계점이 있다면 무엇이라고 생각하나요?
   A. Step size의 과도한 크기 (T = 2000)

6. 본인의 연구에 접목시켜볼 점이 있을지 생각하고 적어보세요.
   A. Diffusion, SR 선행연구

7. 본 Summary를 작성하는 과정에서 생성형AI를 사용했나요?
   A. 아니요

날짜: 2025-08-12

이름: 신준원

Image Super-Resolution via Iterative Refinement

용어 정리

x : Source Image → Conditional DDPM 적용

y : target Image (+ noise 추가)

model 목표: $P(y|x)$ 근사 → direction

역방향 (Reverse process): $p_\theta(y_{t-1}|y_t, x)$

↳ (x, y_0)

SR3 Architecture

Similar to DDPM.

① U-Net Base

② ResBlock (from BigGAN)

③ Re-scale skip Connection ⇒ $\frac{1}{\sqrt{2}}$

④ Block 개수 증가

⑤ bicubic interpolation

⇒ low resolution Image → target과 동일한 resolution 맞추는 interpolation

→ low resolution (x size는 target과 동일) + target

C (channel concat) ⇒ ex) B C H W → B, 2C, H, W

(시간 → FiLM for Conditioning. 단순 Concat Best)

⑥ $T = 2000$

⑦ Variance (uniform sampling) → speech.

* 개념잡기.

Regression Based 인 경우.

$y = Hx + n$

low → high. noise.

Down-sampling

ex) bicubic interpolation // $f(y) = x$.

↳ $f(\cdot)$ 구하기

y 에 대한 f 를 구하기 Loss : 줄이기.

Regression ex) SRCNN.

Generative based ?

↳ 사후평균 (mmse) 을 구하는 distribution 구하기.

↳ $\int x\, p(x|y)\, dx$.

high. / low 분포 · x ⇒ x_{pm} ⇒ posterial mean ⇒ Regression based

그러나 Generative

↳ posterial 구하기

어려움.

Experiments

Dataset

① Flickr-Faces-HQ (training) ⎫
② CelebA-HQ (evaluation) ⎭ → for Face (uncondition)

③ ImageNet 1k → for natural Image super-resolution (condition)

Step) Image → bicubic interpolation. (with anti-aliasing)

Step) 64x64 → {256x256, 512x512}

16x16 → {128x128}

256x256 → {1024x1024}

Result

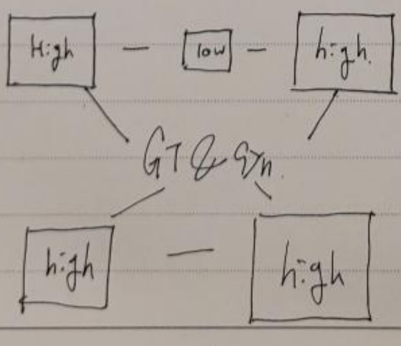
기존의 MOS (mean opinion Score)

단점 → resolution 반영X / 둘 중 한개 선택하는 구조

SR3에서 사용한 방법

step1)

step2)



Q) 어떤 high가 더 나은가요? (low에서 생성된)

Q) Camera에서 찍은 이미지는 둘중 어떤 것일까요?

(3초 간보임)

fool rate select model's 선택 → 50 % 수준 → 왜 low와 함께 ⇒ 판단 비율이 높은가?

↓

Imagenet ⇒ 차이 ×.

저자의 추측 → image 2장 순서로 → 판단이 더 쉬움.

Cascaded generation

장점 ⇒ 곧바로 Generate 보다 자원 ↓

예시 64²(input) → model1(확대) 256² → model2(확대) 1024² → 1024².