Paper summary

AI VISION Lab

1. 공부한 논문의 제목, 게재된 학회 혹은 저널 등 논문 기본 정보를 적으세요.
   - **A. 이름: Trustworthy Super-Resolution of Multispectral Sentinel-2 Imagery With Latent Diffusion**
   - **B. 저널: JSTARS**
   - **C. 도메인: SR, RS**
   - **D. 출판연도: 2025**
   - **E. 저자:**

2. 논문에서 제안한 알고리즘 및 프레임워크에 대해 본인이 이해한대로 다이어그램을 그려보세요. **논문 Figure를 그대로 따라 그리면 안됩니다.**
   A. AE의 경우 LDM의 AE와 동일함.

   B. 이때, 차이점이라 할 수 있는 점은 Wasserstein Distance (바서 스타인)을 구해, Regularization을 적용한 부분이라고 할 수 있음.

   C. 논문에서는 해당 Regularization을 통해 Latent space 정보 유지, data Distribution의 반영 효율 증가 등을 주장함.

   D. Decoder의 경우는 GAN의 Decoder를 가져왔는데, 이때 Condition을 주입하기 위해 backbone(embedding)을 가져옴. 해당 backbone은 3채널로 학습된 구조다 보니, 본 논문(Target: 4bands)에서는 4채널에서 3채널을 Random하게 선택 -> 추출하는 구조로 해결함

   E. Loss의 경우, Regularization + MAE + GAN(Decoder)로 구성되었으며 해당 Loss의 계수는 Hyper parameter임.

   F. Denoiser (SR) 부분의 경우 AE의 Output을 기반으로 Diffusion step을 진행 (이때, 과정은 다음과 같음.
      i. Branch 1: 앞서 학습한 Enc에 (HR + Noise) 주입
      ii. Branch 2: 앞서 학습한 Enc에 (Interpolated LR) 주입
      iii. 두개의 Branch Output을 Condition으로 Diffusion 주입

iv. 해당 과정의 경우는 L2 (HR, Output)을 통해 학습함
v. 이후, 앞서 학습한 Dec를 통해 SR Output 획득 (Sampling과정 – DDIM)

G. Trustworthy의 경우, Hallucination, Omission 등을 고려해서 수치로 판단. 또한, CI: Confidence Interval 함수를 제안함. 이를 통해 Output의 pixel 값 기반 분포를 통해 Uncertainty 분포를 제시.

3. 본인이 생각하는 이 논문의 장점이 무엇이라고 생각하나요? **논문 Contribution bullet을 그대로 따라 적으면 안됩니다.**
   A. R, G, B 뿐만 아니라 NIR의 SR까지 고려한 점.

   B. Trustworthy를 고려 -> Uncertainty를 판단하기 위한 다양한 metrics을 사용한 기준점을 제시했다는 점.

4. 이 논문을 읽으면서 느낀 점, 혹은 배운 점이 있으면 적어보세요.
   A. SAR에서 Robust Feature를 효과적으로 추출하기 위해서 -> ViT혹은 Backbone(ResNet, VGG 등)을 load해서 사용하고 싶었으나 이는 3band만을 가지고 학습되었다는 한계가 있었는데, 이를 해결하는 과정을 확인할 수 있었음.

5. 이 논문의 한계점이 있다면 무엇이라고 생각하나요?
   A. LDM을 접목시켜, AE과정까지 성능은 우수했던 반면, GAN에서 파생된 Decoder의 경우, AE 대비 성능이 떨어짐을 확인할 수 있었음 (본 논문의 Figure에서 -> AE와 Decoder(output)의 PSNR, SSIM을 비교해본 결과, AE의 성능이 output 대비 항상 높은 것을 확인할 수 있었음=equal line보다 항상 아래에 분포하는 모습 확인가능)

   B. 즉, 본 논문에서는 R, G, B, NIR 총 4개의 Band SR을 구성했다는 기여점만 있을 뿐, 실제 구성을 위해 노력했던 Decoder의 경우 알맞지 않다고 판단됨.

   C. NIR까지 고려한 새로운 Task라는 관점에서 기존 SOTA모델이 존재하지 않음. 따라서 비교할 모델의 상대적인 양이 부족했음

6. 본인의 연구에 접목시켜볼 점이 있을지 생각하고 적어보세요.

A. Feature를 추출하는 Backbone의 사용법은 적용할 수 있다고 판단됨. 해당 논문에서는, Feature 추출을 VGG기반으로 진행했으나, 일반적인 backbone은 R, G, B만으로 학습되었다는 한계 때문에, Multi-Spectral한 RS Image와 알맞지 않는다는 한계를 가짐. 해당 논문에서는 이를 해결하고자 Random하게 4개의 Band에서 (R, G, B, NIR) 3개를 선택하는 방식으로 Feature를 추출함.

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

날짜: 2025-08-20

이름: 신준원

Trustworthy Super-Resolution of Multispectral Sentinel-2 Imagery with Latent Diffusion

Introduction

high resolution Remote-sensing Image ←→ low resolution
⇒ 한계점.

GAP ⇓ ▷ Contribution

Cross-sensor super-resolution 으로 해결!

⊕ trustworthy model (Application 측면 ⇒ 예상 결과) — "LDSR-S2"
goal : 2.5m (from 10m)
Architecture : Latent diffusion based.
trustworthy ⇒ uncertainty metrics 평가 기반 (Probabistic model base)

Related work

기존의 SR → RGB 위주 (multi-spectral SR 부족)
스펙트럼 ⇒ NIR Band SR ? ⇒ 공간적 흐림.
⊕ S2 Dataset model

training strategy

XBand 사용 (RGB & NIR)
LR-HR pair 사용

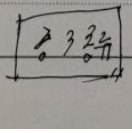
사용 Dataset

① WorldStrat ▷ SPOT 6/7 (HR) + S2 (LR)
- 12 000 장 (가로×세로)
- Spectral 정보 포함.

실험 Dataset

OpenImages → (256) 가로×세로 (500k 장)
LR (bicubic interpolation 생성)
+
Gaussian blur

10m 40m → interpolation 사용
⊕ S2 (HR) + S2 (LR)
⊕ SEN2NAIP 사용



Architecture

LDM (Latent Diffusion model base)

① Auto-encoder

input image (512×512×4) → AE → z (128×128×4)

일반적인 LDM의 encoder와 차이점.

⇒ wasserstein Distance (WD) regularization (AKA. Earth-mover) (바서스타인)

↳ AE (model) ⟷ Prior distribution (HR Image)
 ↳ regularizing 역할 ✓

⇒ AE, prior(HR) → Distance 작음 = 좋음!
 ↳ 닮지 않을수록 / 다를수록 더 큰 규제

↳ goal : Convergence

* metric = Distance.

ex) U / Hilbert / (L_2).

* 왜 washerstein Distance인가?

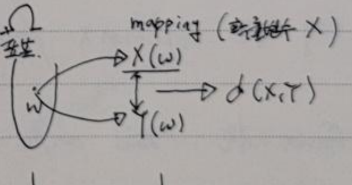
(WGAN ⇒ 사용)

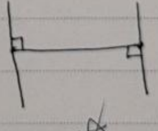
① Latent space 정보 유지
② data distribution 반영 ↑.
③ 효과적인 regularization

* Wassenstein Distance 정의

Joint $\pi(H, Q)$에 대해서.

$d(X, Y)$의 기댓값을 최소화하게!





⇒ 거리값 최소.

② Decoder. (GAN)

- Lpips, VGG ⇒ 3Bands 구조.
 ↳ 해당 논문에서는 4Band 中 3Band를 random 선택하여 train 진행.) 고려해볼만한 접근방식.

③ Loss

$$L_{total} = \lambda_{WD} \cdot L_{WD}(z) + \lambda_{MAE} \cdot L_{MAE}(\hat{x}_{HR}) + \lambda_{GAN} \cdot L_{GAN}(\hat{x}_{HR}) + \lambda_{Lpips} L_{Lpips}(\cdot)$$

④ Denoiser

(해상도 키우기)

Step1) Low-resolution x_{LR} → "Interpolate"

Step2) Interpolated LR into Encoder (Latent) ⇒ z

Step3) Encoding x_{HR} 시작.

Step4) Encoded x_{HR} ($= A_{enc}(x_{HR})$) 에 Diffusing
(이때, reverse ⇒ $p_\theta(x_{t-1} | x_t, z)$)
↳ encoded LR.

Step5) Loss = L_2. (L_1 보다 sampling 자연스러움)
↳ (Saharia et al.)

⑤ Sampling (DDIM Based)

LR → interpolation = LR' ~~$A(LR') = z$~~

※ 이때 추가 Condition X ⇒ ~~$\frac{RGB-NIR}{original}$~~ 에 Condition으로 넣음

Diffusion 거친 output x_0 → refine 과정 (with Decoder)

Result

Robust ⇒ Decoder가 decoder 하기 쉬워짐.

* 본 실험에서는
AE 대비 output(SR)의 성능을 비교.

→ SSIM & PSNR.

→ 결국 AE 이후 SR의 역할 ↓.

⇒ 즉, 해상 능력의 Decoder 성능 ⇒ AE output 성능을 대변하는 역할을 보장.

다양한 metric 적용

- Reflectance Consistency : RMSE (LR ↔ Down sampled HR)
- Spectral Consistency : Spectral angle (spectral fidelity)
- Spatial Consistency : Spatial alignment
- Synthesis : Spatial resolution (quantify)
- Omission : HR ↔ SR
- Improvement : 향상
- Hallucination : 환각

신뢰성을 평가.

↳ function CI

↳ Confidence interval

⇒ CI(x) = [$\bar{x}$ - ϵ, $\bar{x}$ + ϵ]

↳ Gaussian distribution 가정.