

라이언로켓 인공지능연구원 사전과제

모든 과제는 nothing 보다 something 에 의미를 둡니다. 😊
전부 완료하시지 못하더라도, 결과물을 제출하여 주신다면
최선을 다해 긍정적으로 평가하겠습니다.

과제는 총 이론과제와, 구현과제 두 가지로 구성되어 있습니다.
각각 다른 역량을 평가하기 위한 것이오니 한 가지에서 많이
부족하다 느끼시더라도 크게 걱정하실 필요 없습니다.

제출기한은 다음 주 수요일 9/1 일 오후 6 시까지입니다.
결과물을 parkjs767@lionrocekt.ai 로 메일 전송 부탁드립니다.

1. 이론 과제

1. 다음 논문 중 한 편을 리뷰하시면 됩니다.

2. 리뷰 시 아래의 다섯 가지 질문에 대해 한 줄 요약과
해당 논문에 대한 의문점, 한계 등 개인의 생각을
PPT 2 장 정도로 정리해서 제출해주시면 됩니다.

▪ 질문

1. 배경: 문제를 제기하게 된 배경
2. 문제: 저자가 해결하고자 하는 문제
3. 해결방안: 문제 해결을 위해 도입한 방법론
4. 근거: 해결방안을 도입한 계기, 근거
5. 기여점: 이전까지 논문들과의 차이점, 기여점, 개선사항

▪ 논문

1. Glow-TTS: A Generative Flow for Text-to-Speech via Monotonic Alignment Search, Kim et al., 2020.

<https://arxiv.org/abs/2005.11129>

2. Is normalization indispensable for training deep neural networks? Shao et al., 2020.

<https://papers.nips.cc/paper/2020/file/9b8619251a19057c7f70779273e95aa6-Paper.pdf>

3. HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis, Kong et al., 2020.

<https://arxiv.org/abs/2010.05646>

4. Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech, Kim et al., 2021.

<https://arxiv.org/abs/2107.03298>

5. Non-Attentive Tacotron: Robust and Controllable Neural TTS Synthesis Including Unsupervised Duration Modeling, Shen et al., 2020.

<https://arxiv.org/abs/2010.04301>

6. StarGAN v2: Diverse Image Synthesis for Multiple Domains, Choi et al., 2019.

<https://arxiv.org/abs/1912.01865>

7. Denoising Diffusion Probabilistic Models, Ho et al., 2020.

<https://arxiv.org/abs/2006.11239>

8. First Order Motion Model for Image Generation, Siarohin et al., 2019.

<https://papers.nips.cc/paper/2019/file/31c0b36aef265d9221af80872ceb62f9-Paper.pdf>

9. Multi-label Contrastive Predictive Coding, Song et al., 2020.

<https://arxiv.org/abs/2007.09852>

2. 모델 구현 과제

1. 영상과 음성 둘로 나누어져 있어 희망하시는 분야를 선택해주세요

2. 분야별 세 가지 중 하나를 선택해 모델을 구현하시면 됩니다.

- 영상분야 지원자

1. 비디오 또는 이미지 합성, 생성
2. 비디오 또는 이미지 분류, 객체 인식
3. 기타 확률 및 생성 모델

- 음성분야 지원자

1. 음성 합성, 변환 (Text-to-Speech, Voice-conversion)
2. 음성 인식
3. 기타 확률 및 생성 모델

3. 모델 구현이 처음이시라면 아래 가이드라인을 참고해주세요

- 구현 모델의 선정

1. 영상분야 : resnet bottleneck block 고도화 모델(resnext 등), visual transformer, image preprocessing pipeline 등
2. 음성분야 : Tacotron2, FastSpeech, Glow-TTS 등

- 구현 방식

1. python tensorflow, pytorch, jax/flax 환경 권장 (pytorch 추천) 그 외 버전 및 개발 환경은 무관
2. GPU 환경을 가정하며 기기가 없는 경우 Colab 혹은 캐글을 활용
3. 그 외는 자율적으로 구현

- 제출 방식

1. github/colab/kaggle 등을 통해 코드를 제출할 것
2. 생성된 음성 혹은 이미지 샘플은 샘플 페이지 링크를 공유하거나 압축하여 제출
3. 학습한 checkpoint는 github release 혹은 압축파일을 통해 제출
4. learning curve는 readme 혹은 ppt 형식으로 제출
5. 합성하는 방법, 개발환경의 경우 readme 혹은 ppt 형식으로 제출(파이썬의 경우 requirements.txt 권장)
6. 오픈 소스를 참고해도 되며, 참고할 경우 링크와 라이선스를 리드미 혹은 ppt에 함께 첨부