Part 1. Disentanglement and Compositionality in Computer Vision

- Disentangled Representation Learning은 ML model이 feature space를 semantic meaning을 갖는 disjoint factor들로 나눠 이해하도록 학습하는 방식임

- DRL은 model의 explanability 외에도 task에 맞는 representation만 분리 가능한 generalizability와, generation 시에 factor를 조작할 수 있는 controllability도 향상시킬 수 있음

- DRL method는 우선 representation structure에 따라 dimension-wise DRL과 vector-wise DRL로 분류할 수 있음

- dimension-wise method는 각 generative factor를 latent vector Z의 한 개의 차원으로 표현하며, VAE와 GAN model 구조에 의존하는 초기 DRL 연구에서 자주 사용됨

- 또한 fine-grained factor를 갖는 synthetic, simple dataset에만 적용할 수 있음

- vector-wise method는 각 factor를 하나의 latent vector로 표현하기 때문에, 높은 정보량의 coarse-grained factor로 이루어진 복잡한 real-world dataset에 적용되고, 최근 더 주목받고 있는 방식임

-DRL method를 model type에 따라 VAE, GAN, diffusion based method로 나눌 수 있음

- VAE-based model 중 dimension-wise method인 beta-VAE는 KL divergence loss 항에 beta coefficient를 추가해, beta가 클 때 disentangled representation 학습에 높은 가중치를 두게 됨

- 또한 reconstruction과 disentanglement 간 trade-off를 고려해 KL term에 information capacity를 학습에 따라 늘어나도록 한 method 이외에도, 기존 VAE에 regularizer를 추가한 여러 변형 method들이 제안됨

- GAN-based model dimension-wise DRL method인 InfoGAN은 생성자 input에 noise z와 함께 latent variable c를 넣고, 생성자가 c의 정보를 잃지 않도록 mutual information term I(c;G(z,c))를 추가함

- 더 나아가, InfoGAN-CR은 InfoGAN에 contrastive regularizer를 추가해 생성자에서 샘플링한 image들로부터 fixed dimension을 찾아내도록 했고, PS-SG GAN은 각 latent dimension의 focused area를 가리키는 spatial mask를 학습함

- DR-GAN은 vector-wise DRL을 적용했고, pose-invariant face recognition을 위해 identity와 pose를 각각 나타내는 두 개의 vector를 사용함

- diffusion-based model 또한 제안되었는데, 그 중 하나인 Disenbooth는 denoising 과정에 identity-preserved part, identity-irrelevant part로 분류하는 disentanglement task를 추가함

- 지금까지 언급한 DRL methods는 generative factor들을 parallel하게 가정하는 flat DRL method지만, hierarchical DRL은 factor를 계층적 abstraction level로 나눠 분류함

- FineGAN은 hierarchical level마다 하나의 latent representation을 사용해 background, parent, child code로 나눠 fine-grained object generation을 수행함

-