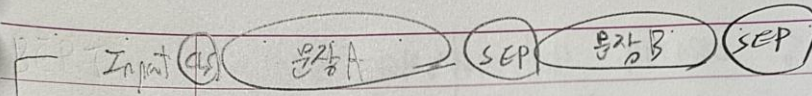
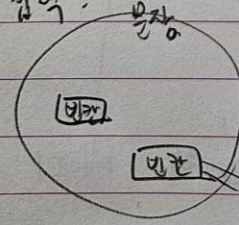


# BERT

PAGE  
DATE



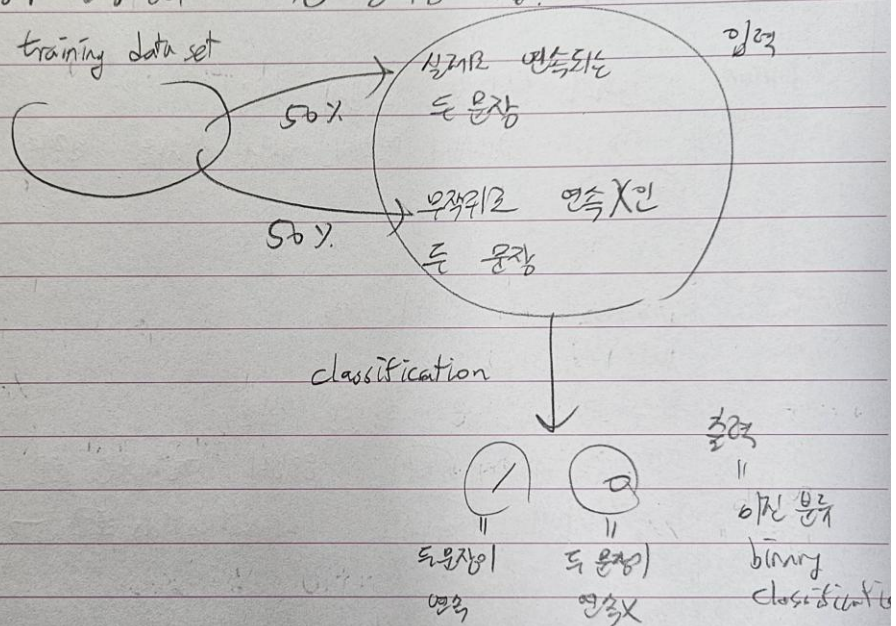
Token 임베딩	Pre trained 워드 임베딩
Segment 임베딩	이력 또는 문장   문장 A에 속하는지 / 문장 B에 속하는지 알려줄 역할. <u>항상 대입</u>
Position 임베딩	각어의 순서 정보를 모델에 제공. 언어가 시퀀스에서 몇 번째 위치에 있는지를 알려줌. <u>항상 대입</u>
Bert Idea	Transformer 의 인코더 사용. Self-supervised. 매우 큰 dataset을 활용한 학습. 문장 사이에 [SEP] : 문장의 경계 구분 토큰. 문장 시작에 [CLS] : classification 토큰 → 이후 분류 작업에 활용
Bert training 1.	MLM = Masked Language Modeling 입력 : 문장 출력 : Masked 토큰을 가려진 단어를 맞추는 것.  Masked 토큰으로 가림.
	Loss function 최소화 : 원래 문장 단어 (x) 예측 문장 단어 (y) → 비교해서 Loss function 계산



대형 코난	통의역 등 여러개의 양이 차승하라는 은리
활동	→ 이런 noise에도 불구하고 ;
	모텔이 문맹상 정결한 양이들을 들은 활동은 예슬

bert . NSP = Next Sentence Prediction.

training 2. 문장과 문장 간의 거시적인 관계를 학습



7월 2주. self-supervised learning

≠ Herd | labelling  $\frac{1}{2}$   $\frac{1}{2}$   $\frac{1}{2}$   $\frac{1}{2}$

= 원시 데이터 / 928 data 만 있으면,

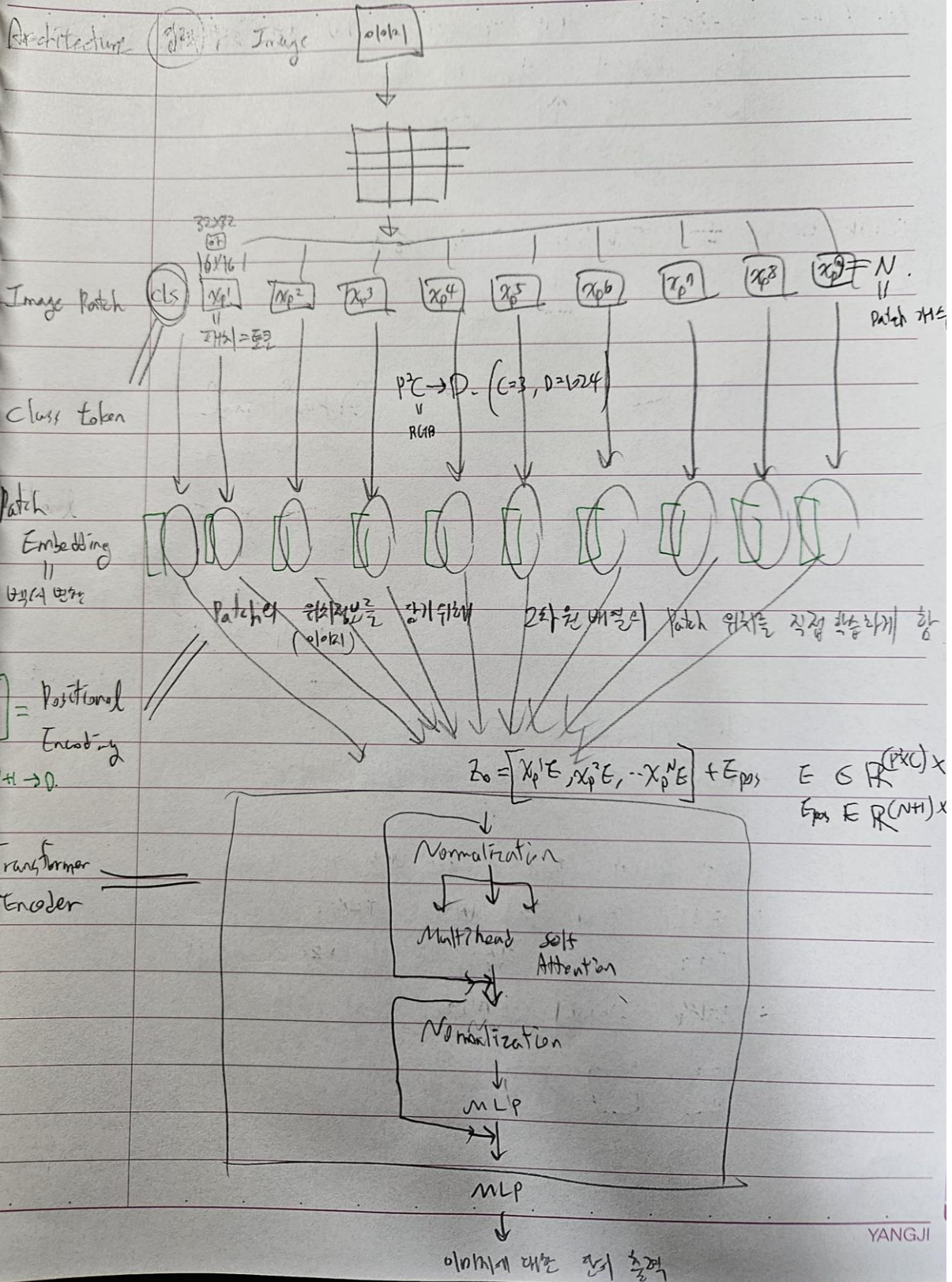
Mark 원치 / 연속 여부 등 2로 00 후 재입력 생략 가능

= 3046: min & nsp

= 위의 training으로 BERT를 워드 임베딩 한다.

→ 이는 오랫동안 워드영에 많이 표출







상 / 비유 2020년도 가장 강력했던 CNN 차원 모델 ResNet-152

보다 정확도가 약간 더 높게 나옴.

but 비유 ↑↑ 여건함. → 비효율적  
시간 ↑↑

VIT가 비효율적 = Inductive bias를 사용 X

비유 이유

||  
CNN에 있음  
그치지

① Spatial  
Locality

||

본래 인접을 위해  
주변 픽셀을 참조함.

but, ViT는 차원  
임.

③ Positional  
Invariance.

||

이미지의 모든 위치의  
동일한 필터 적용

but ViT는 이런 공간적 지역성 X  
이미지 전체 정보로

→ 함께 Patch Embedding 활용

결론

데이터로부터 모든 것을 학습, ViT는.

→ CNN이 갖는 전체적인 문맥 정보도 파악 가능.

Positional  
embedding  
활용하기

입점한 위치의 Patch들이 서로 비슷한 값을 가지며,  
별어질수록 값이 점차 변하는 패턴을

VIT가 스스로 학습할 수 있는 것을 나타냄.

→ 2차원 공간이나 위치관계 명확히 파악

ex)

