

# 자기지도학습 특강

세상에 쓸모없는 데이터는 없다

진행자: 신해솔, 김언지

튜터: 문이안, 박현영

# Content

## 1. Background

- 1-1 supervised learning
- 1-2 latent vector
- 1-3 data and labeling

## 2. Self supervised learning

- 2-1 signal
- 2-2 Self supervised learning
- 2-3 pretext task
- 2-4 downstream task

## 3. Examples

- 3-1 Rotation Task
- 3-2 SimCLR
- 3-3 MAE

## 4. Conclusion

## 5. Code practice

**Background**

# Supervised learning

지도학습



고양이

label

강아지

training



고양이

/

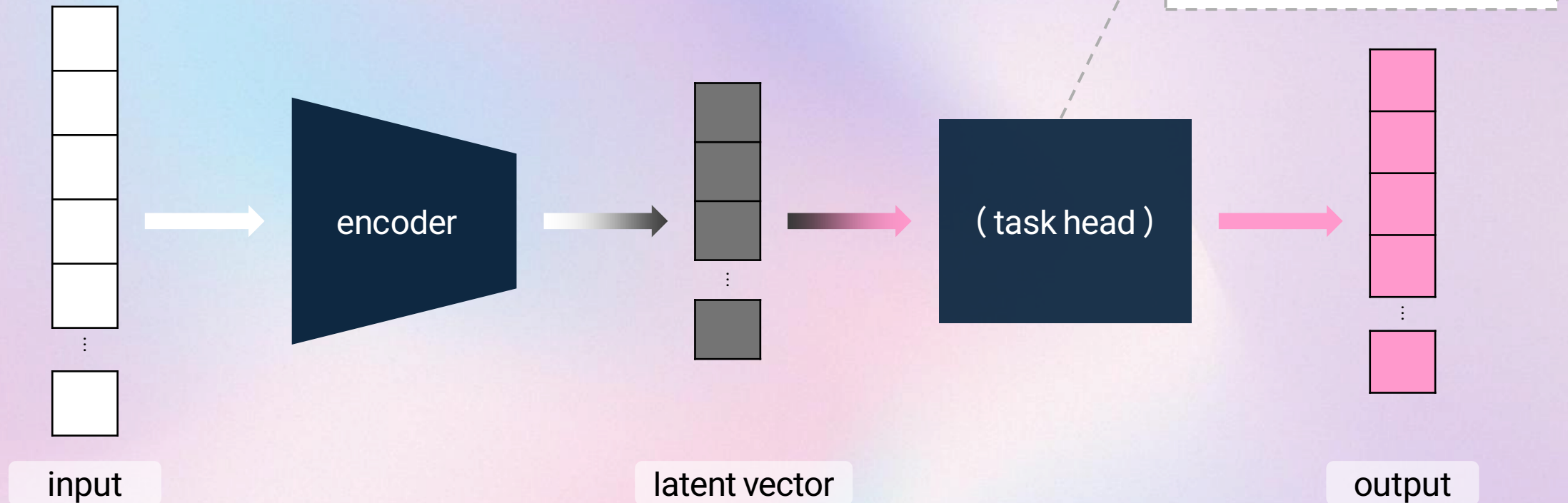
강아지

데이터 + 레이블 -> 모델



# Supervised learning

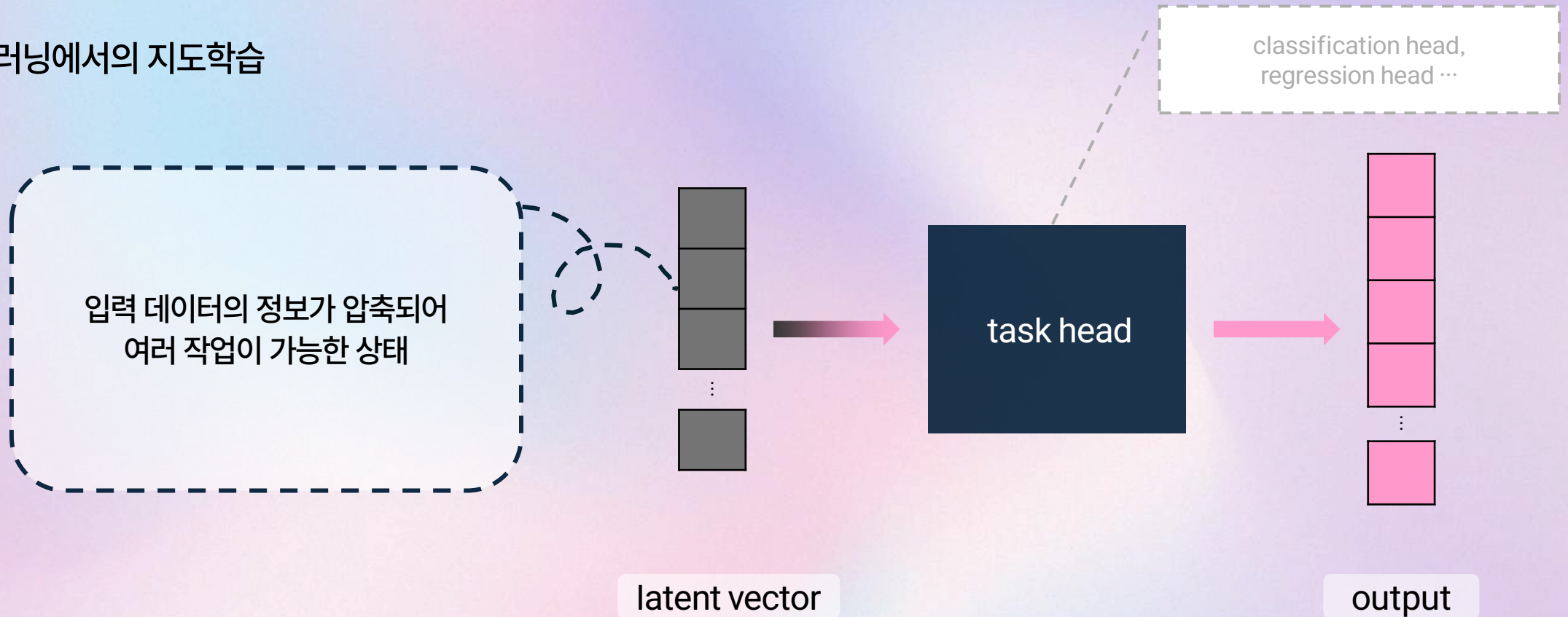
딥러닝에서의 지도학습



입력 데이터를 처리하는 과정에서 필연적으로 중간 단계의 벡터를 생성하게 됨

# Supervised learning

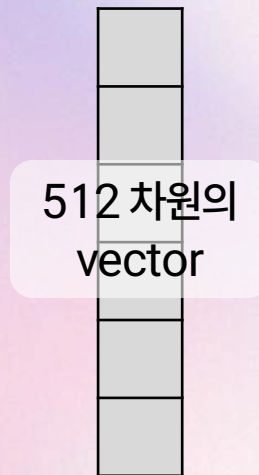
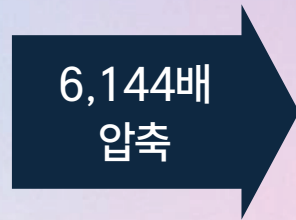
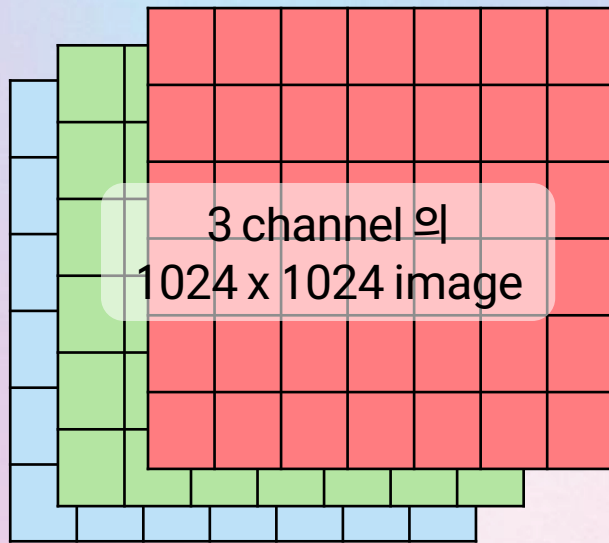
딥러닝에서의 지도학습



입력 데이터를 처리하는 과정에서 필연적으로 중간 단계의 벡터를 생성하게 됨 = latent vector

# Latent vector

왜 latent vector를 만들어 사용하려는걸까?



계산의 효율성

노이즈 제거

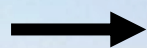
일반화 능력

고차원 데이터의 **본질을 저차원으로 압축!**

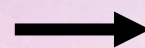
# Latent vector

즉, LATENT VECTOR는!

고차원 데이터의 저차원 압축 표현 : 원본 데이터의 핵심 특성을 담고있는 것



encoder



latent  
vector

0.62

0.98



이 벡터  
안에  
나 있다.

여기서 고차원 데이터 = 이미지, 텍스트 등

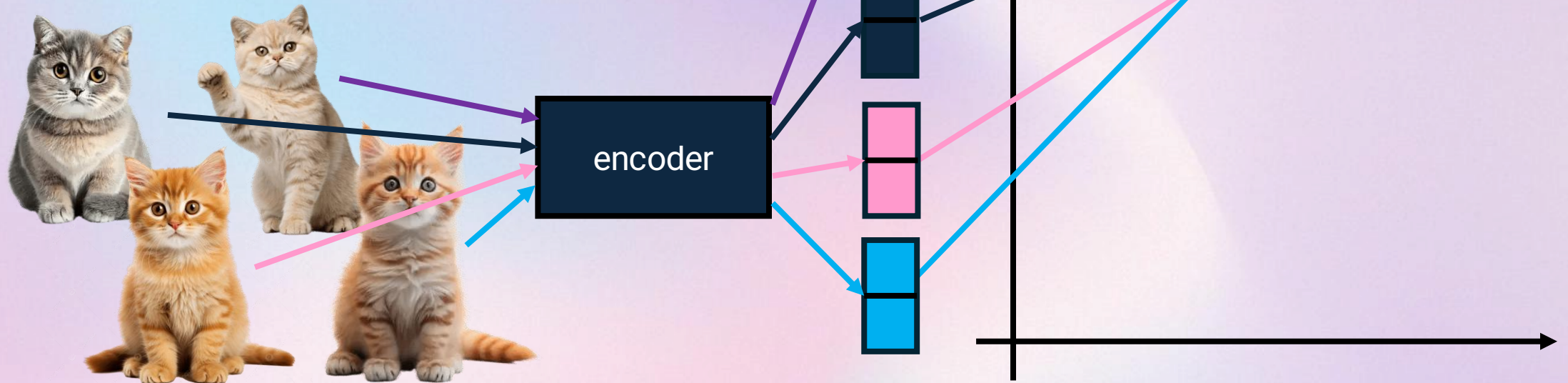
여기서 저차원 압축 표현 = 숫자로 표현되어진 벡터



# Latent vector

좋은 LATENT VECTOR?

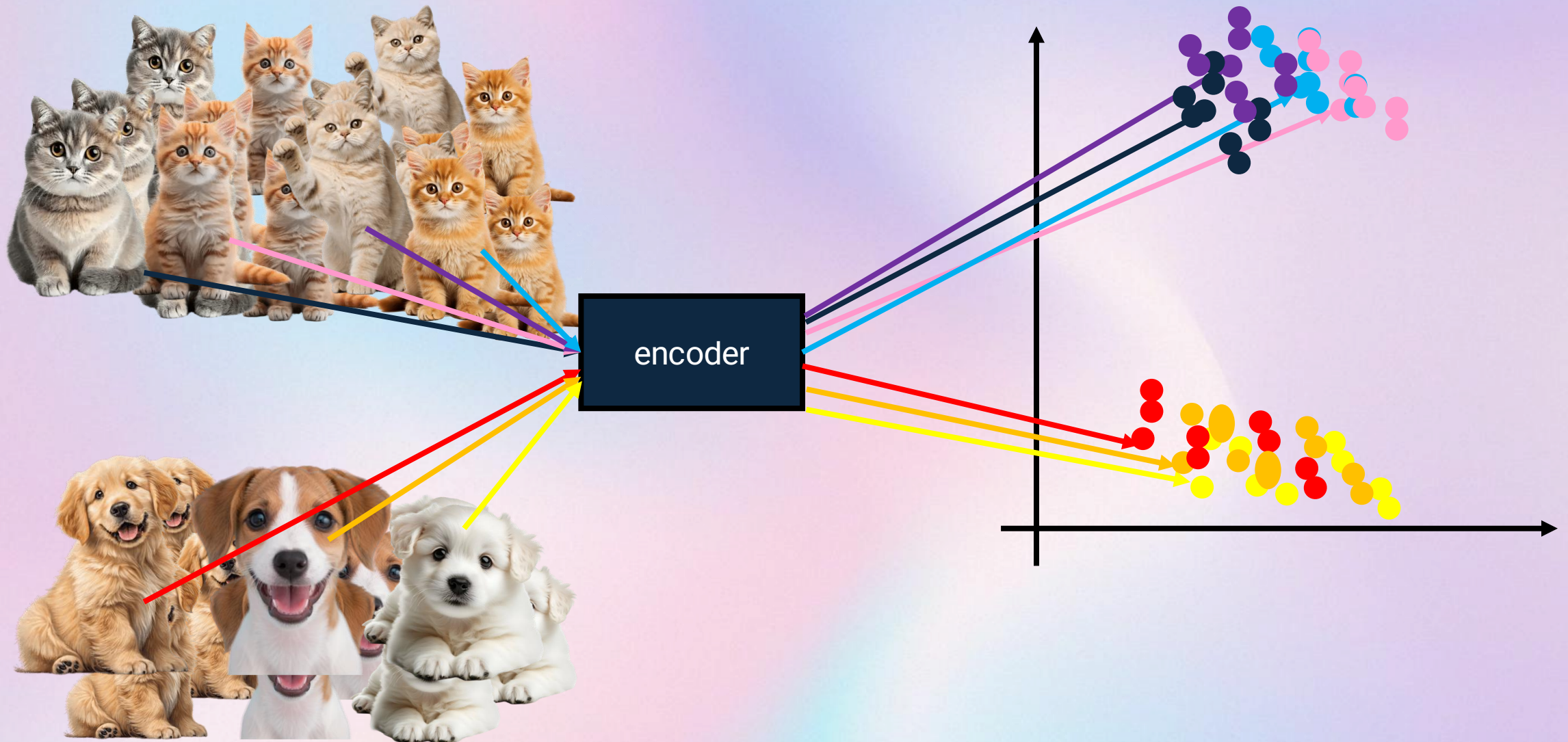
→ 의미적 유사성과 중요한 특성을 담아냄



우리가 모델을 학습시키는 주 목표는

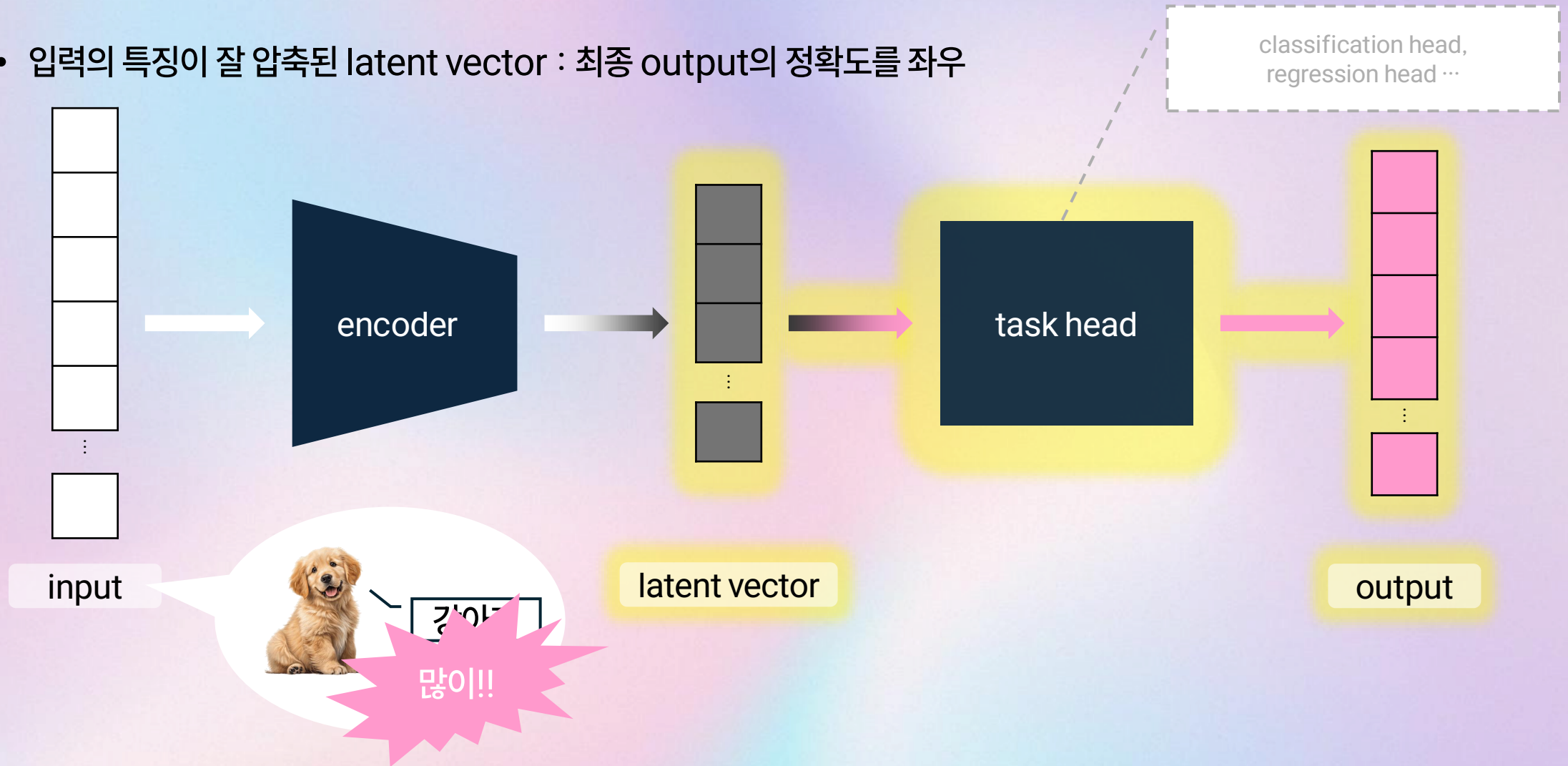
모델(여기서 encoder)이 이 latent vector를 잘 만들어낼 수 있도록 하는것!

# Latent vector



# Latent vector and Learning

- 입력의 특징이 잘 압축된 latent vector : 최종 output의 정확도를 좌우



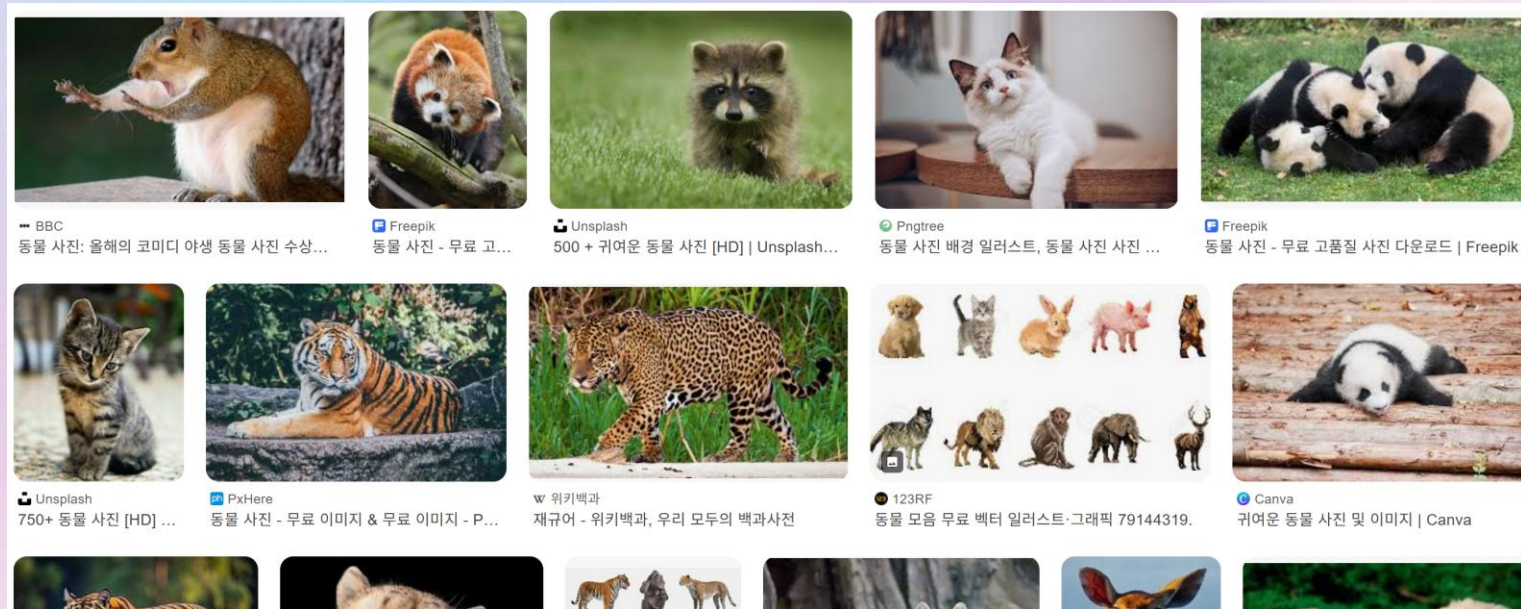


# Data and labeling

라벨링의 현실적 어려움

➔ ImageNet 1400만 장 라벨링에 수년 소요

➔ 주관적 판단, 실수 등의 문제



지도 학습에는  
대량의 데이터 필요

라벨 데이터 부족 →  
지도학습에서의 모델  
성능 제한



# Data and labeling

원본 데이터 자체가 정답이 되어보면 어떨까?



label

고양이



# **Self supervised learning**

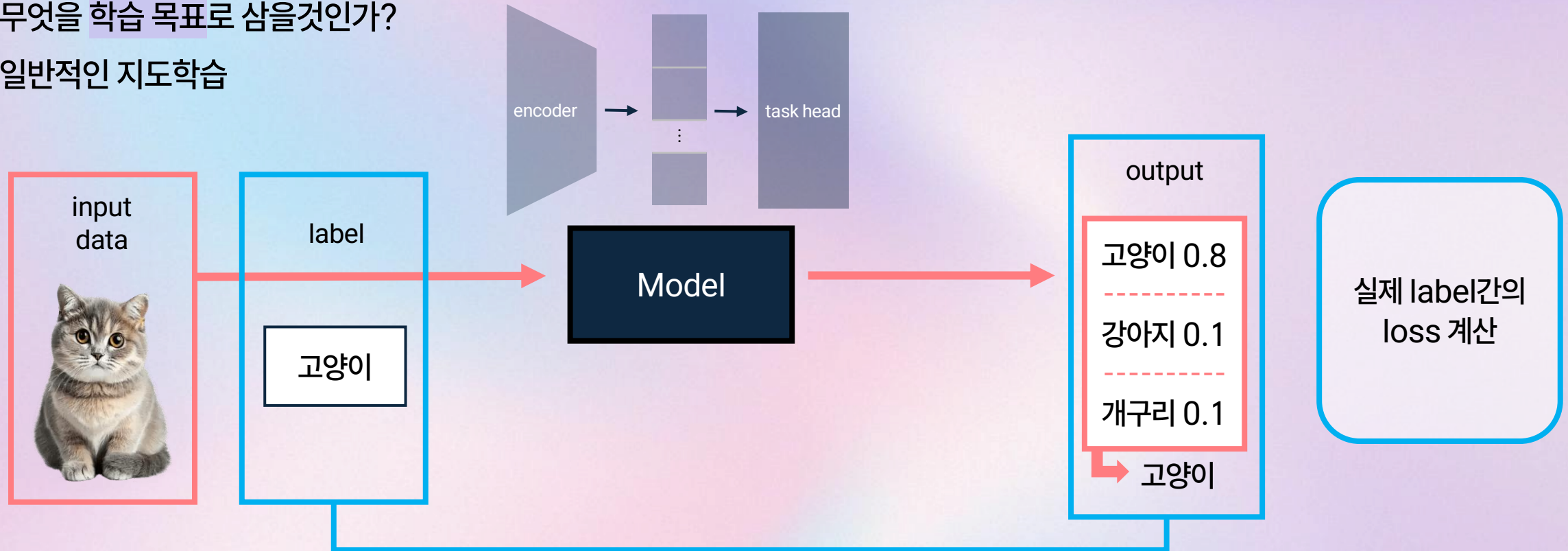
# Self supervised learning

라벨이 없는 데이터에서 데이터 자체의 구조를 이용해  
스스로 감독 시그널(supervision signal)을 생성하여  
표현을 학습하는 방법

# Signal

무엇을 학습 목표로 삼을것인가?

일반적인 지도학습

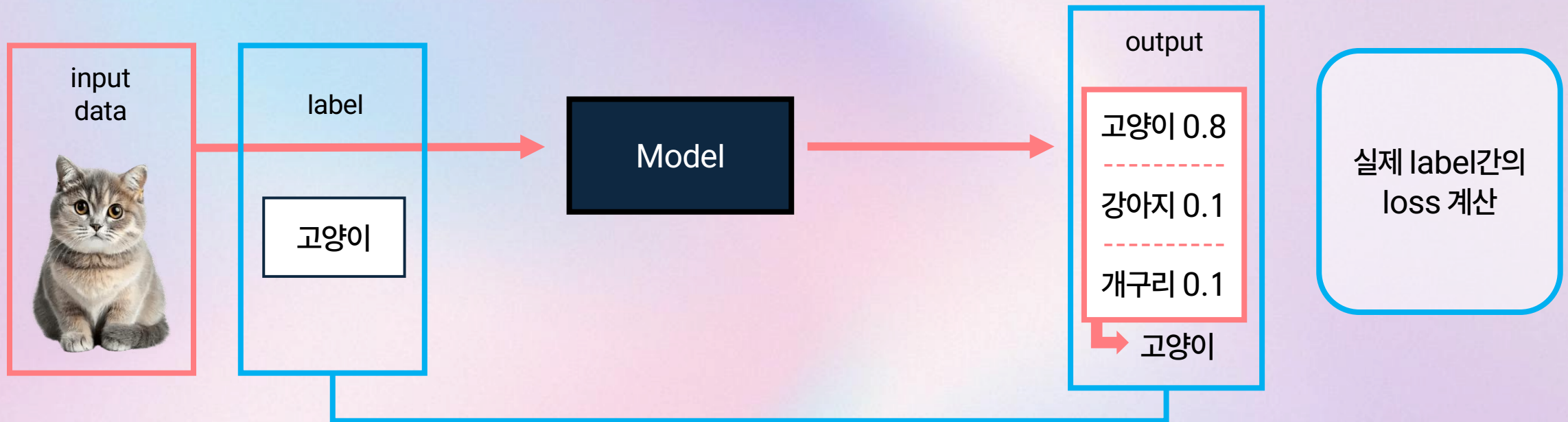




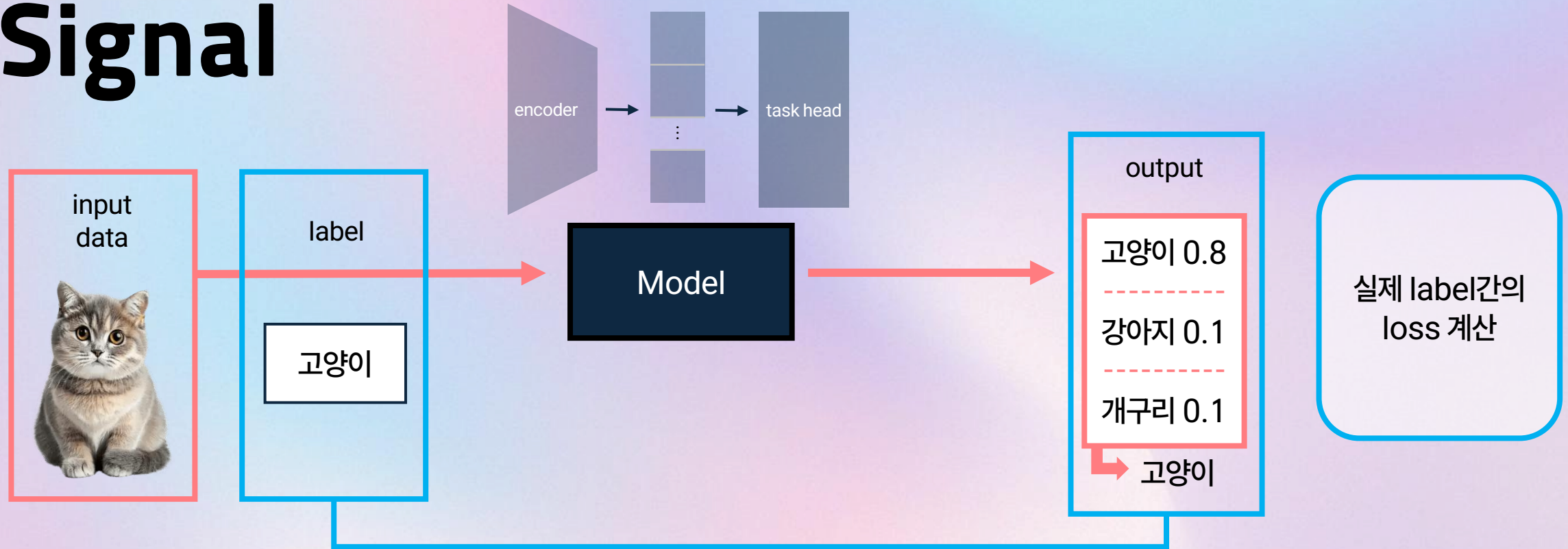
# Signal

무엇을 학습 목표로 삼을것인가?

일반적인 지도학습



# Signal



예측값과 정답 라벨 간의 차이(loss) 최소화

output : [0.8, 0.1, 0.1]

label : [ 1.0, 0.0, 0.0 ]

cross-entropy loss :

$-(1 \times \log(0.8) + 0 \times \log(0.1) + 0 \times \log(0.1))$

$= 0.22$

$$H = - \sum p(x) \log p(x)$$

# Data and labeling




label

고양이



data

A fluffy brown and white kitten sitting down, looking directly at the camera.



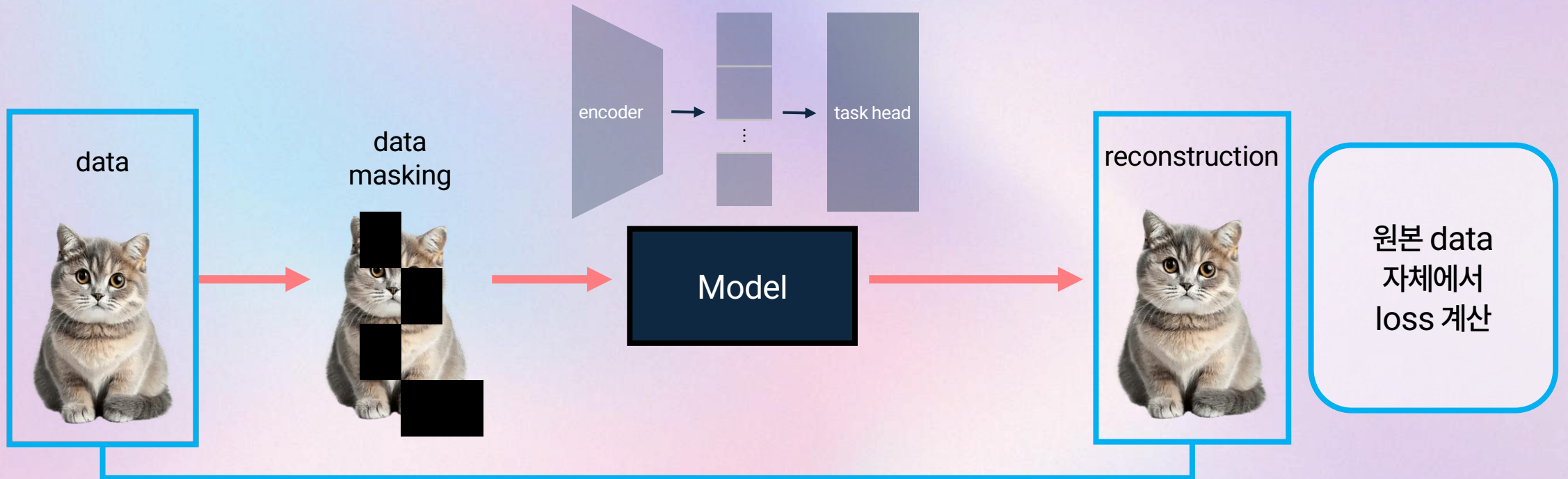
# Self supervised learning

어떻게?

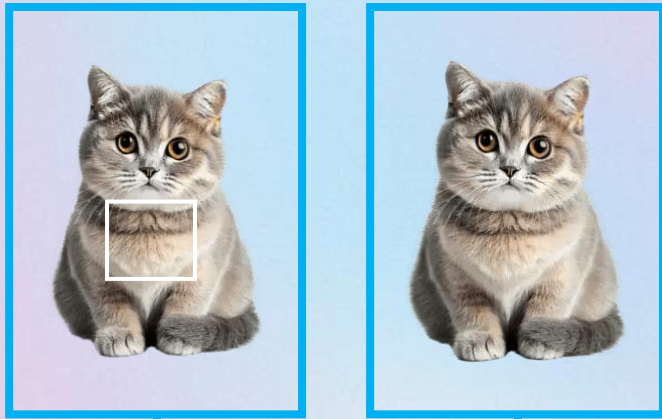




# Self supervised learning



# Self supervised learning



원본 data  
자체에서  
loss 계산

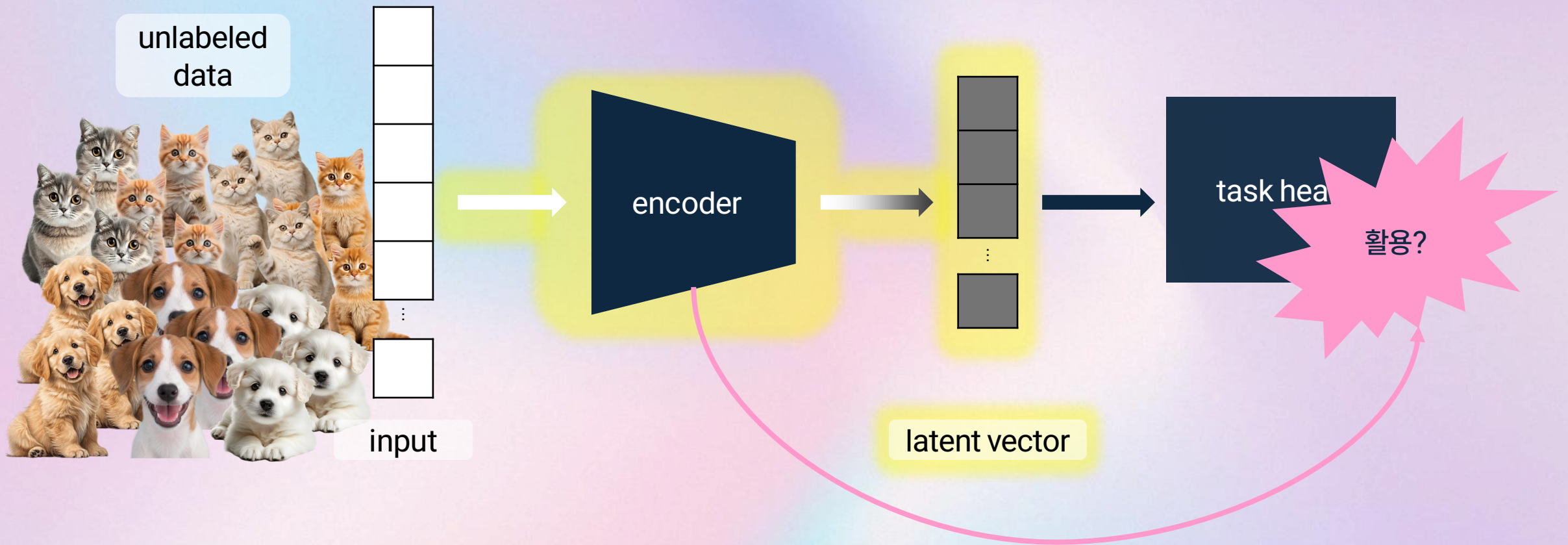
10	20	30	40	50
60	70	80	90	100
110	120	130	140	150
160	150	180	190	200
210	220	230	240	250

12	20	28	38	50
58	68	78	90	98
110	118	12	138	148
158	168	180	188	198
208	218	228	238	250

$$\begin{aligned} \text{mse} &= (4+4+\dots+4+4) / 25 \\ &= 76 / 25 \\ &= 3.04 \end{aligned}$$

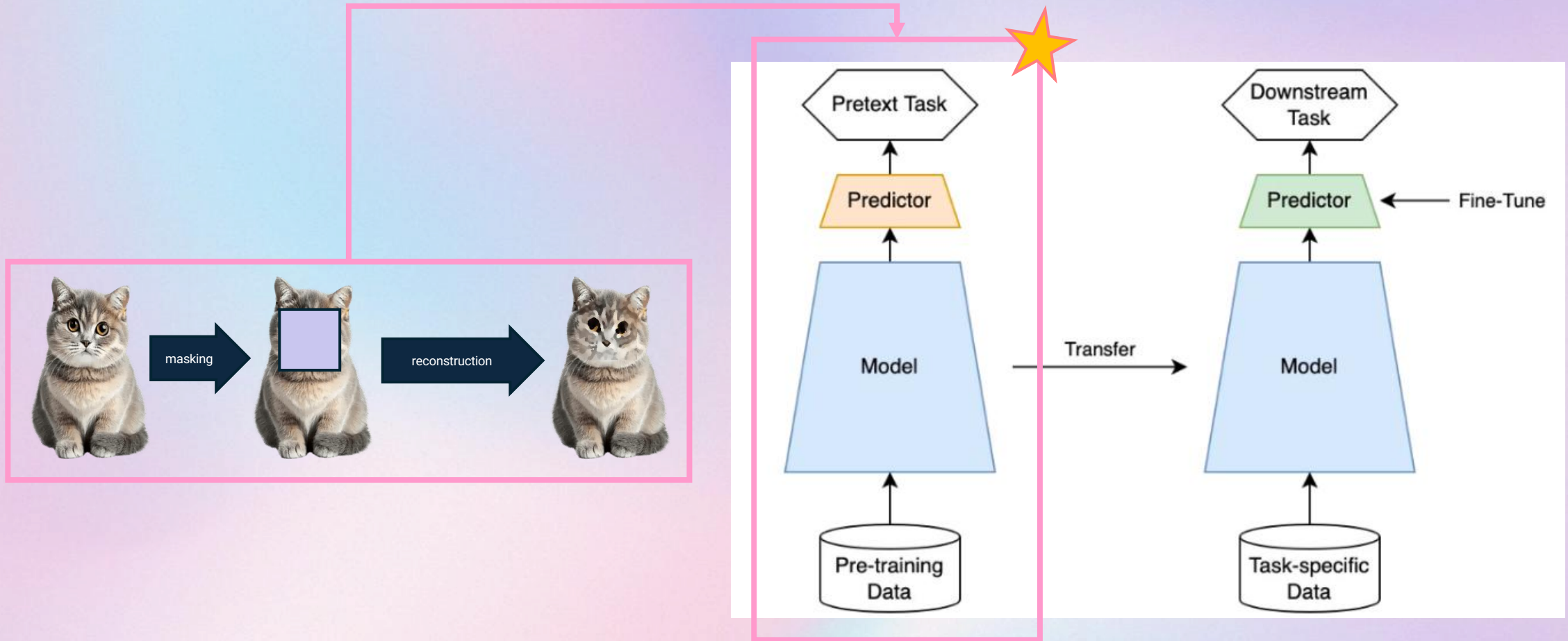
$$MSE = \frac{1}{n} \sum \left( y - \hat{y} \right)^2$$

# Self supervised learning





# Self supervised learning



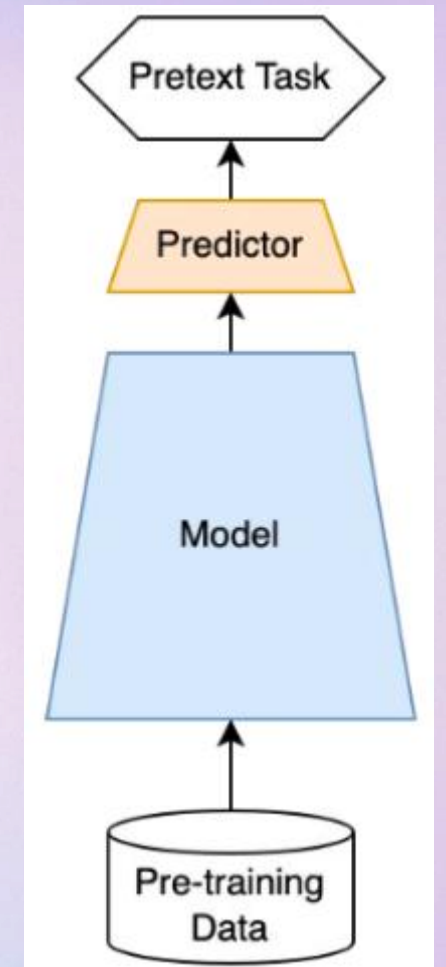


# Pretext task

## Pretext Task

- 데이터에서 자동으로 생성한 보조 과제
- 레이블 없이 데이터의 특징을 학습하는 단계

Pretext Task를 통해 Encoder가 의미 있는 Latent Representation을 학습

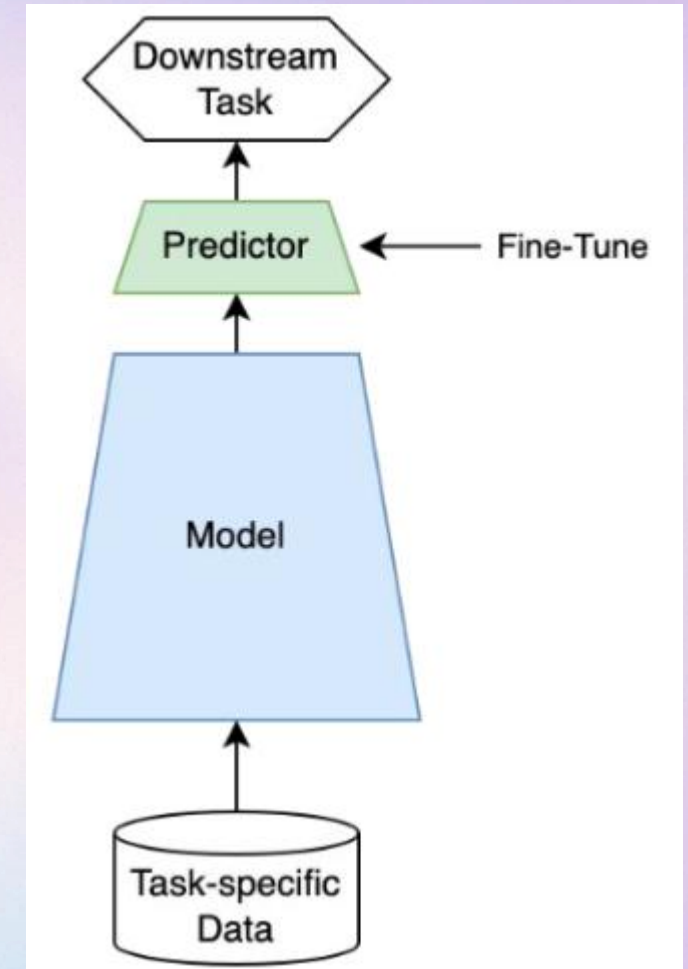


# Downstream task

## Downstream Task

- 실제로 해결하고자 하는 목표 과제
- 분류, 검출, 분할 등 구체적 응용

Pretext에서 학습한 특징을 Downstream에서 활용



# Examples

# Rotation Task

Unsupervised Representation Learning by Predicting Image Rotations

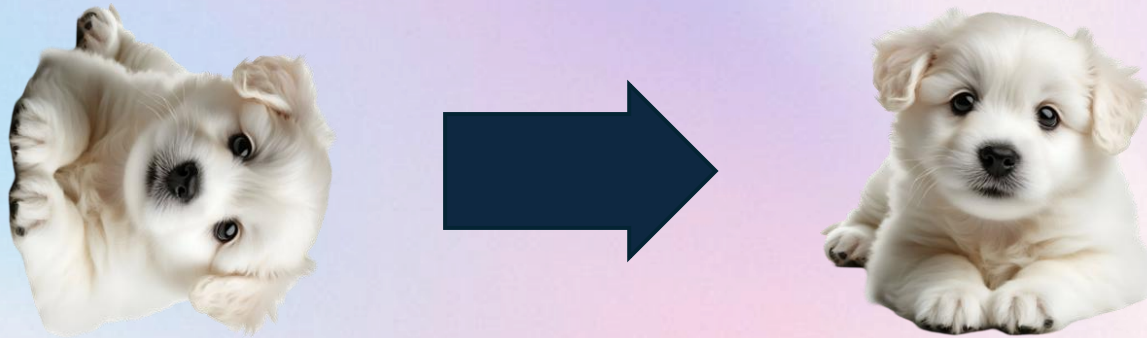


90 rotation



# Rotation Task

Unsupervised Representation Learning by Predicting Image Rotations

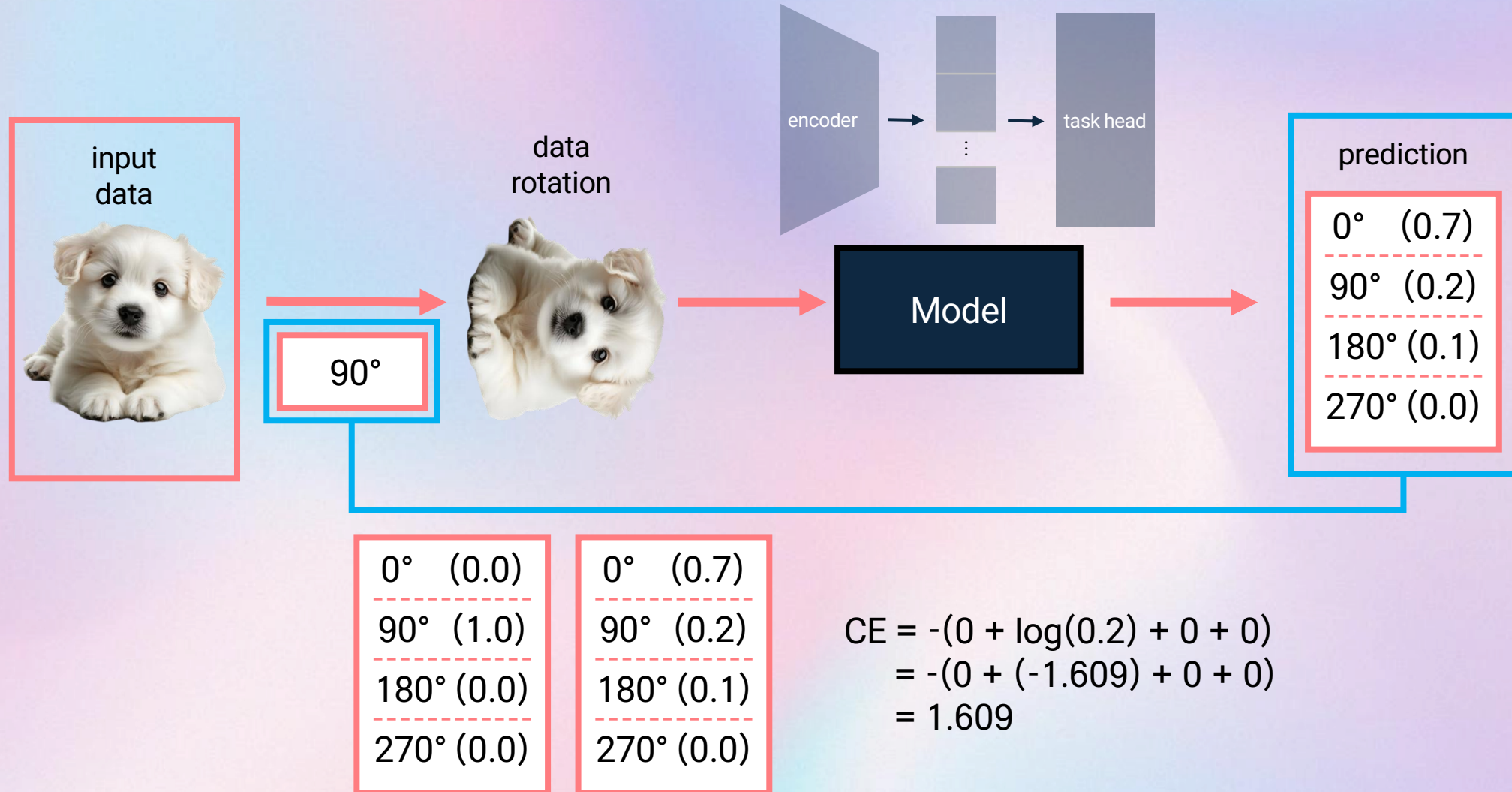


학습 목표 : 이미지가 몇 도 회전했는지에 대한 예측값과 실제 정답 간의 차이(loss) 최소화

원본 이미지를  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ 로 회전시켜 4개의 변형을 만들고,

각 이미지가 몇 도 회전되었는지 구별하며 표현을 학습하는 방식

# Rotation Task



# SimCLR

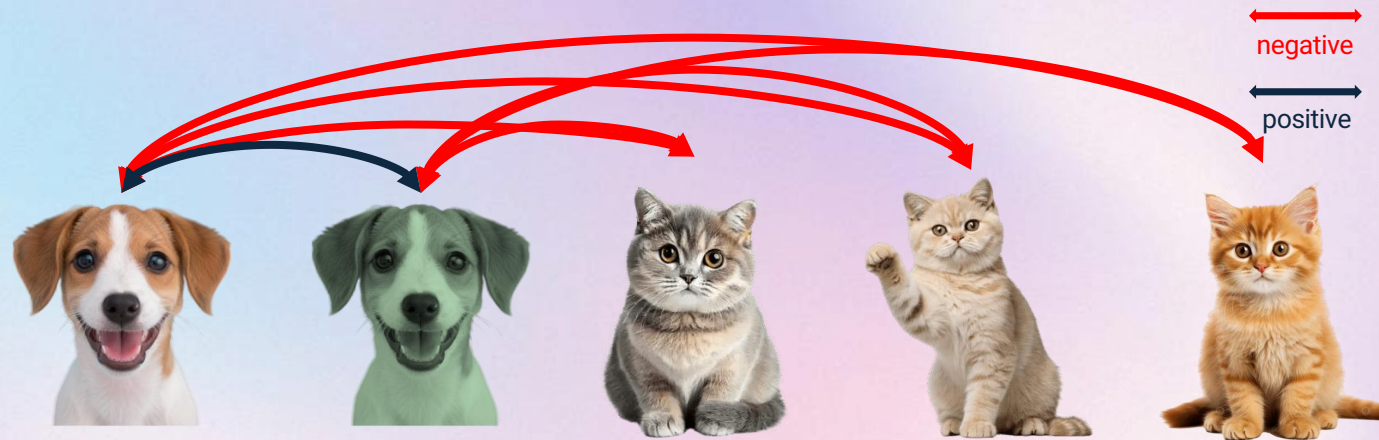
A Simple Framework for Contrastive Learning of Visual Representations





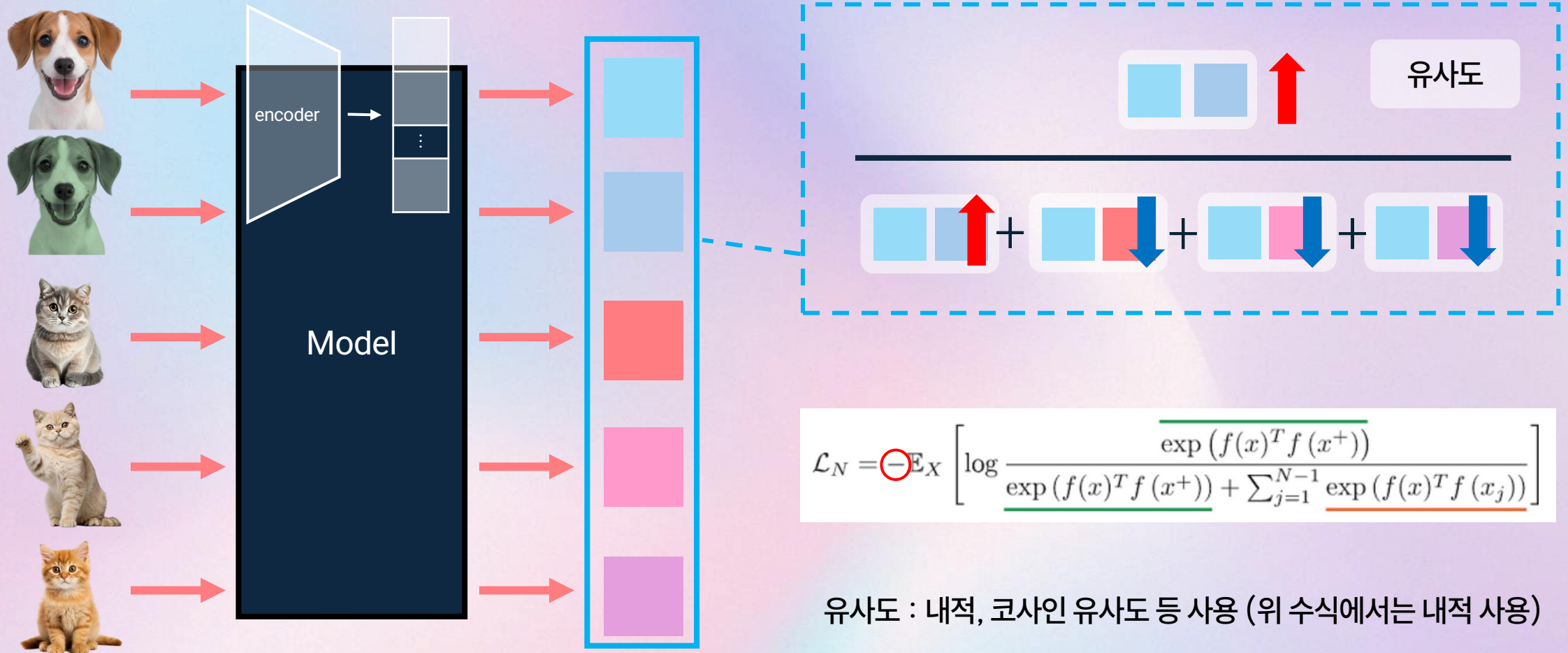
# SimCLR

A Simple Framework for Contrastive Learning of Visual Representations



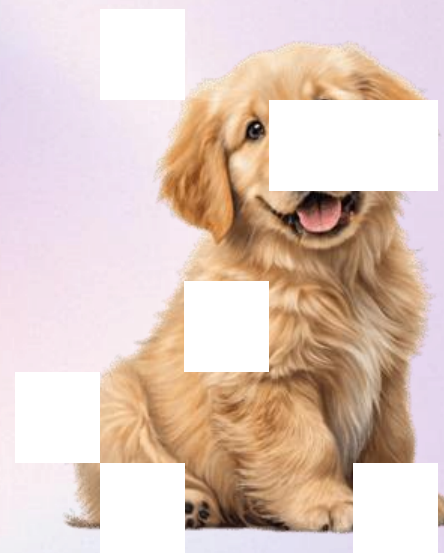
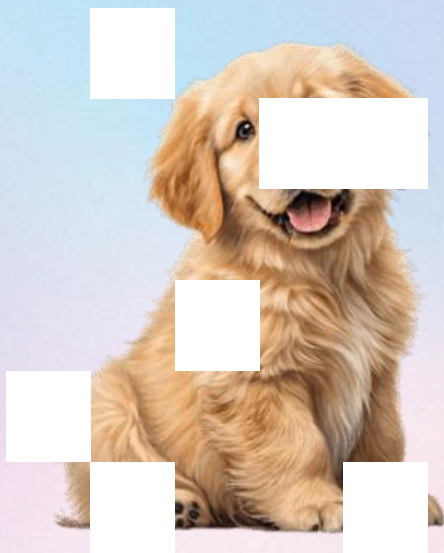
학습 목표 : 두 이미지가 positive pair인지, negative pair인지에 대한 예측과 실제 정답 간의 차이(loss) 최소화  
augmentation된 원본 이미지와 다른 이미지들을  
원본 이미지와 같은 이미지인지 아닌지 구별하며 표현을 학습하는 방식

# SimCLR



# MAE

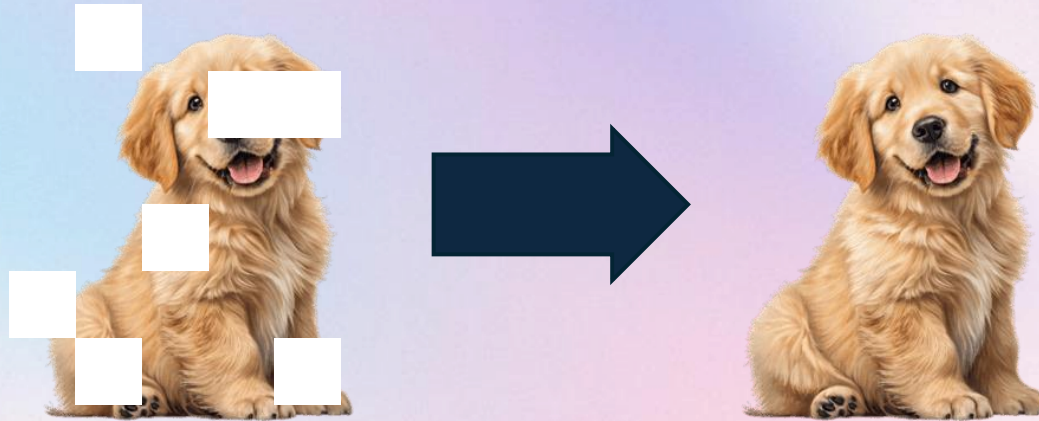
Masked Autoencoders Are Scalable Vision Learners





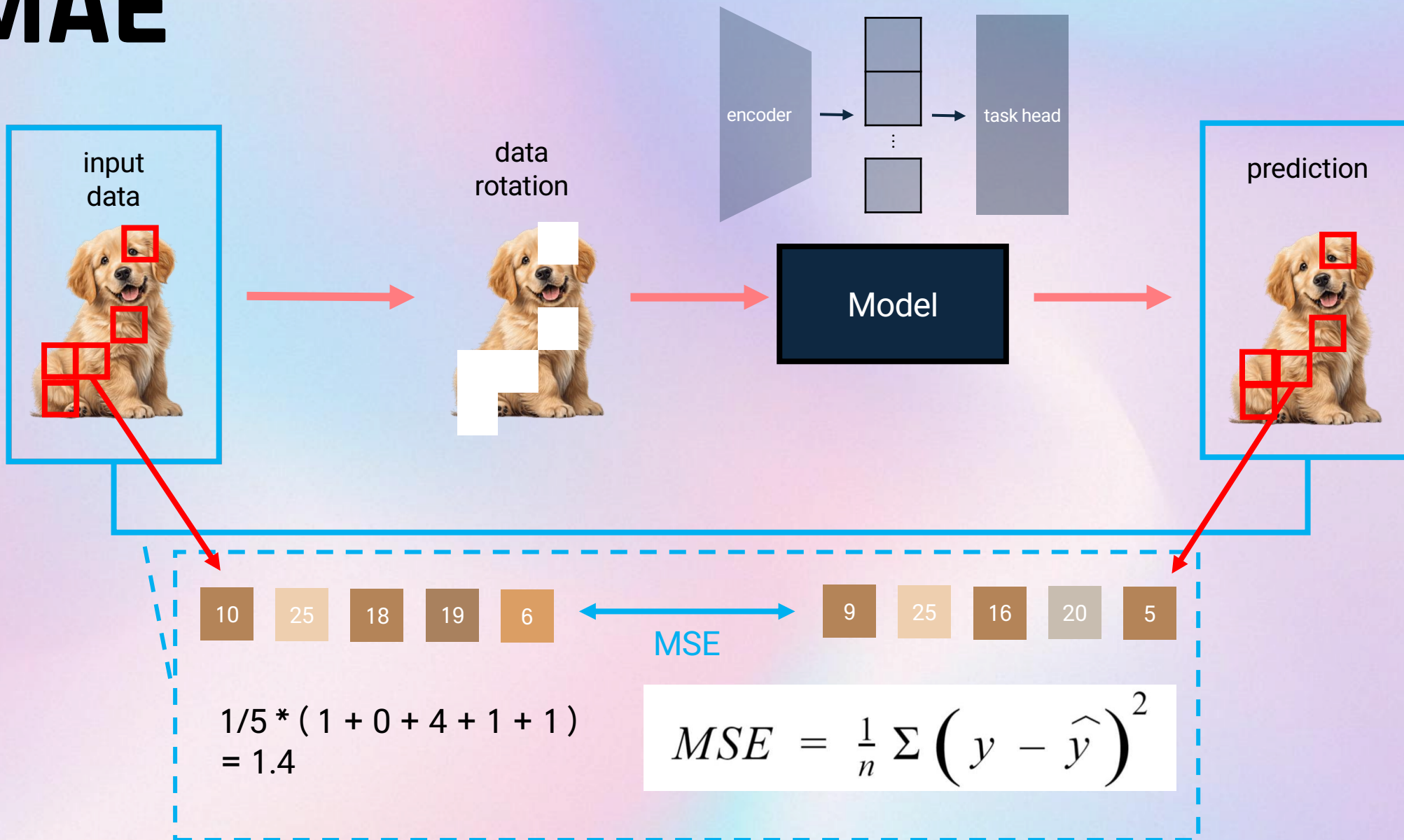
# MAE

Masked Autoencoders Are Scalable Vision Learners



학습 목표 : 마스킹된 이미지의 복원 예측값과 원본 이미지의 픽셀값들과의 차이(loss) 최소화  
이미지의 일부 패치를 랜덤하게 마스킹하고,  
나머지 부분으로부터 가려진 패치의 원본 픽셀값을 재구성해가며 표현을 학습하는 방식

# MAE



# Conclusion



# Conclusion

labeled data는 극히 제한적, unlabeled data는 무한대

➔ 원본 데이터 자체를 정답으로 활용하는 자기지도학습

➔ 지도학습의 가장 큰 문제점인 라벨 데이터 의존성을 해결할 수 있는 핵심적인 해결책

세상은 Unlabeled Data로 가득하다

**code practice**