

인공지능 기술의 대중화(AI Democratization)를 위한

TANGO 커뮤니티 제1회 컨퍼런스



신경망 레이블링 기술

성명 신동민

소속 (주) 웨다



주관



주최



후원



목차

1. 데이터 레이블링

2. 비전 데이터 레이블링

3. 자동 레이블링 기술

4. TANGO 데이터 레이블링

인공지능 기술의 대중화(AI Democratization)를 위한
TANGO 커뮤니티 제1회 컨퍼런스

1. 데이터 레이블링 - 개요

Card News

매경이코노미

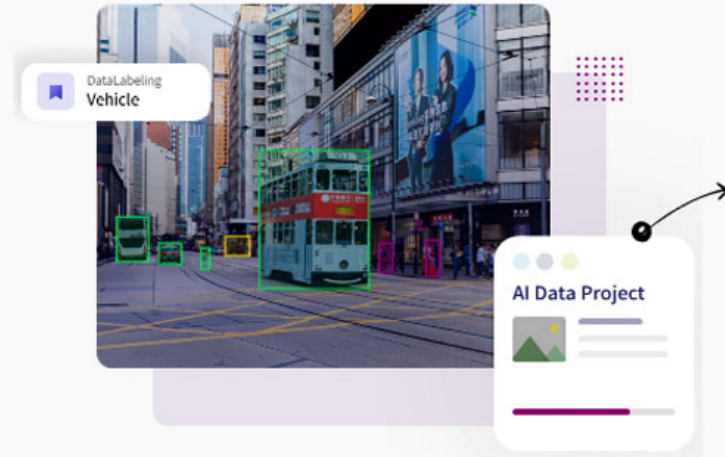
데이터 라벨링이 뭐길래
새롭게 인기 끄는 '디지털 막노동'



Card News

매경이코노미

코로나19 유행 이후 집에서 하는 부업
이른바 '데이터 라벨링'이 인기를 끄니다

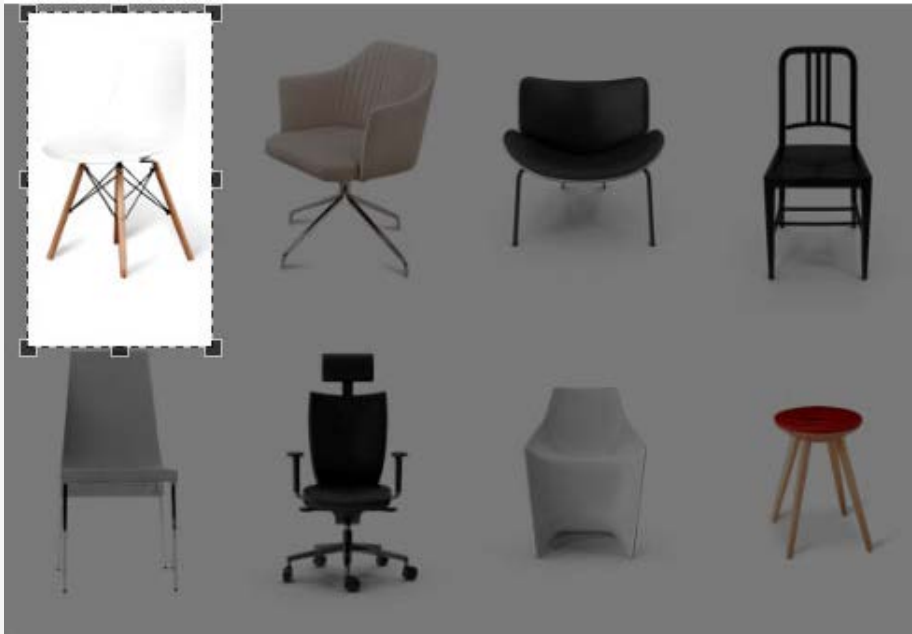


1. 데이터 레이블링 - 개요

데이터 레이블링

인공지능을 위한 학습자료

스스로 배울 수 없는 인공지능의 학습을 위해 정제된 자료를 사람이 직접 모아서 입력하는 과정



1. 아이템을 선택 후 해당 아이템 이미지를 라벨링 하세요.



의자



수정

취소

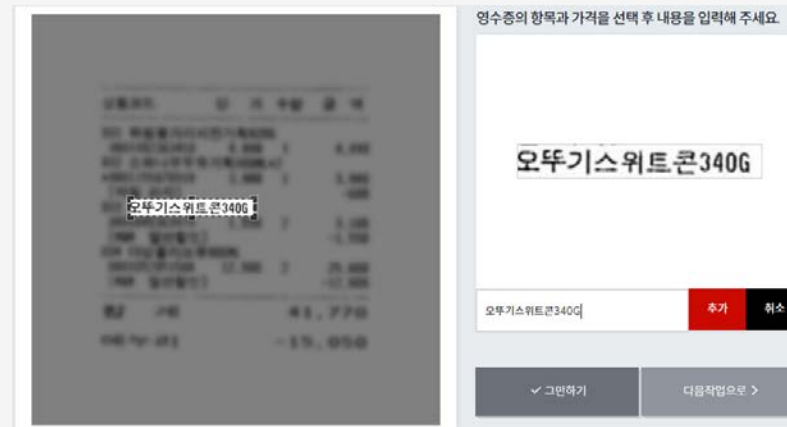
1. 데이터 레이블링 - 개요

수집과 가공

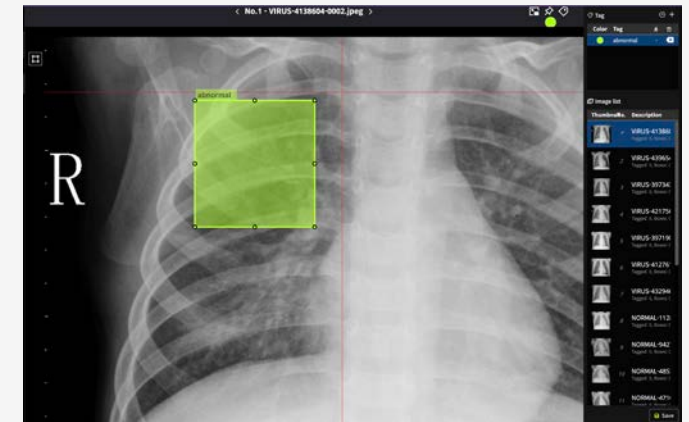
데이터를 모으는 수집과 데이터에서 필요한 부분을 추출하는 가공 단계



음성 데이터 가공



문자 데이터 가공



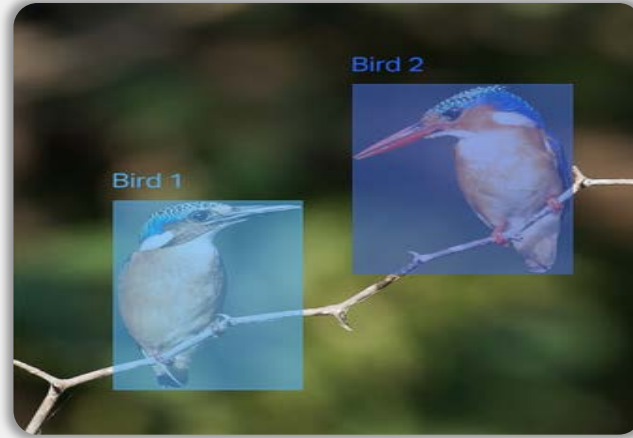
이미지 데이터 가공

2. 비전 데이터 레이블링 - Image Labeling의 종류



Image Classification

- 이미지가 어떤 분류에 속하는지 예측 하는 것



Object Detection

- 여러 물체가 어떤 물체인지 분류 (Classification)
- 물체의 위치 정보를 파악 (Localization)

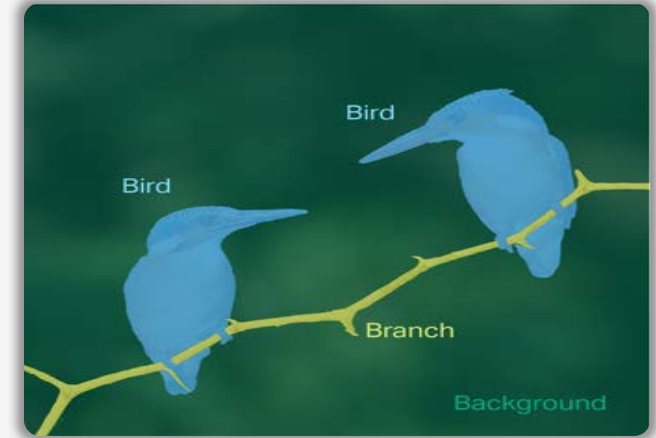


Image Segmentation

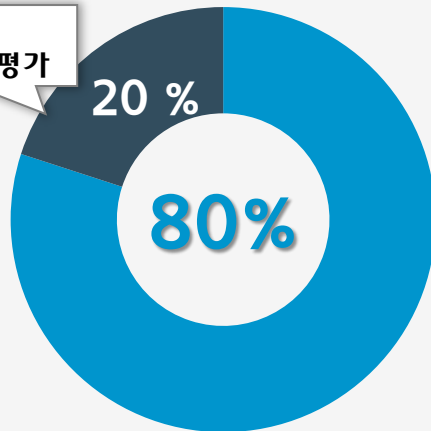
- Pixel을 기반으로 모든 Label을 예측 하는 것

2. 비전 데이터 레이블링 - Human Labeling의 문제점

Image Data 분석 시간

전체 인공지능 개발 과정 중
Data Labeling 이 80% 소요

ML/DL
학습 모델 평가



소요 시간 증가

- AI 학습의 80% 이상을 Labeling이 소요
- 자율주행 영상 1시간 Labeling 작업 시 8시간 소요 (Financial Times)

노동 집약적

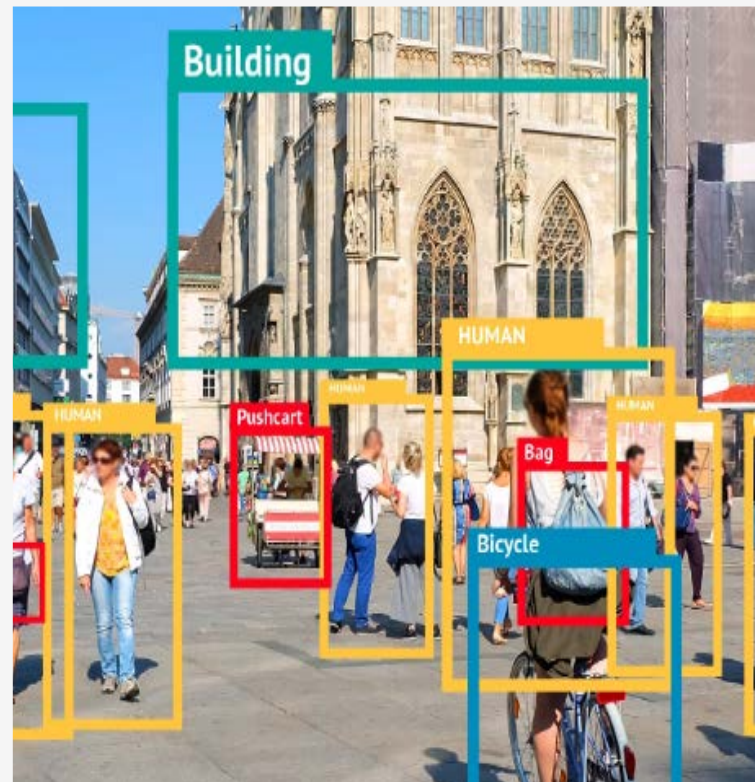
- 사람에 의한 Labeling 과정이 필수

지속적인 데이터 학습

- 데이터 품질이 높을수록 학습 효과가 좋음

Human Labeling 의 문제점

“Auto Labeling 기능을 활용한 사람의 개입 최소화”



많은 사람이 필요



많은 데이터 필요



많은 시간과 비용 소요



품질 저하

3. 자동 레이블링 기술 개요

Seamless한 환경 연계

- 이미지 데이터와 사용자간 데이터 연계
- Labeling, 학습, Auto Labeling 적용 등 GUI 작업 연계

Codeless 분석

- Training 스크립트 코드 작성 여부
- Auto Labeling 실행을 위한 스크립트 코드 작성 여부



사용자간 협업 환경

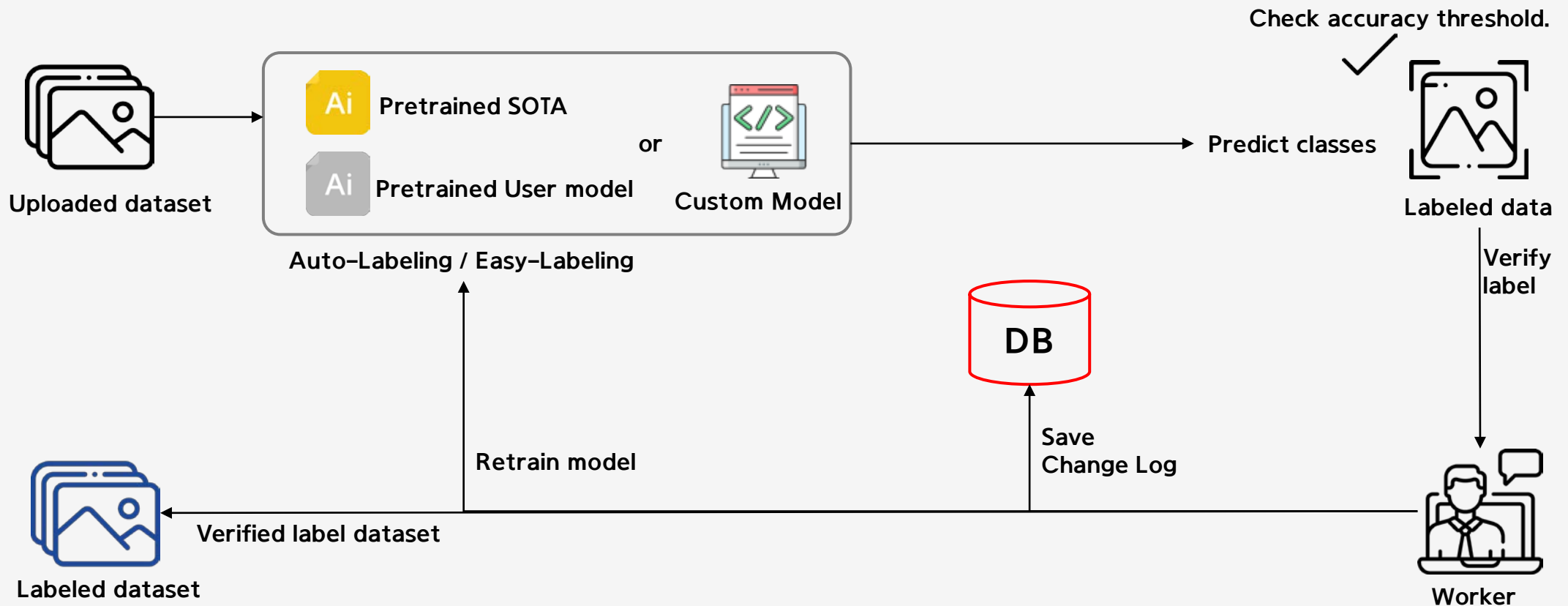
- Dataset 사용자 별 할당 용이 여부
- Labeled Data Cross check 가능 여부
- 분산된 Dataset Merge 학습 가능 여부

모니터링

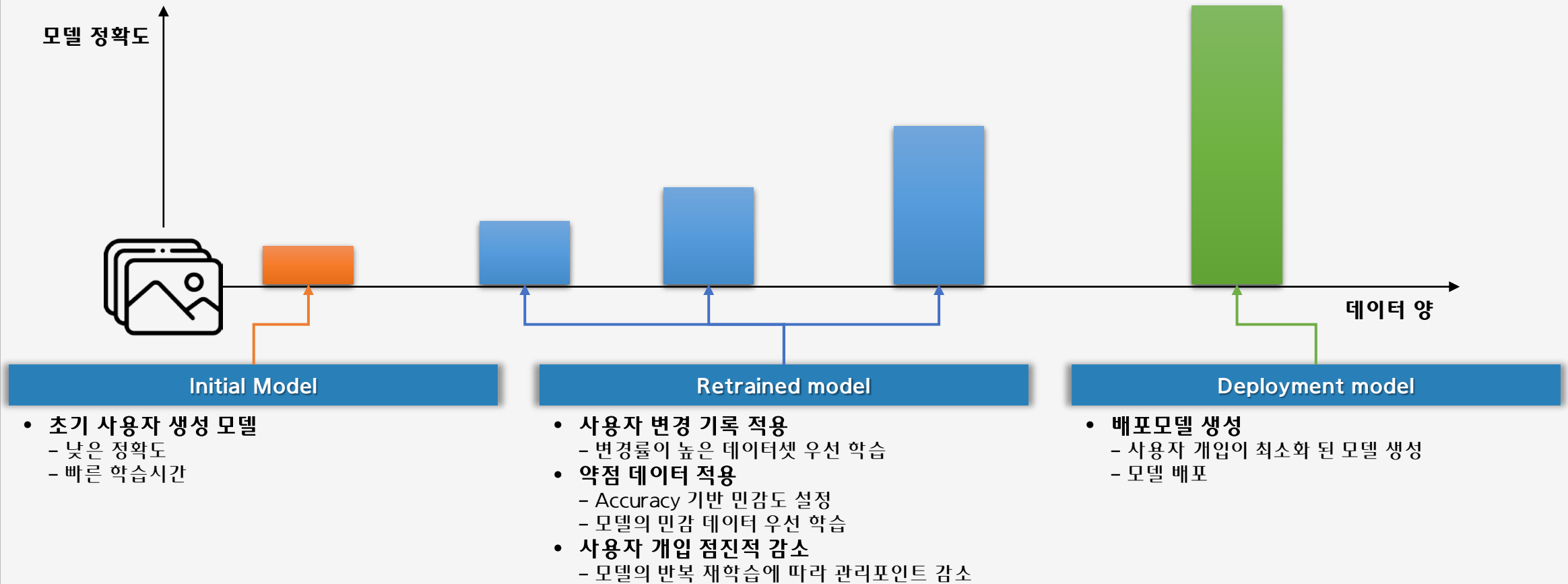
- 모델 학습 진행 상황 모니터링
- 시스템 상황 모니터링
- Auto Labeling 진행 상황 모니터링
- Model accuracy 모니터링

3. 자동 레이블링 기술 - 프로세스

Auto Labeling 프로세스



3. 자동 레이블링 기술- 기대효과



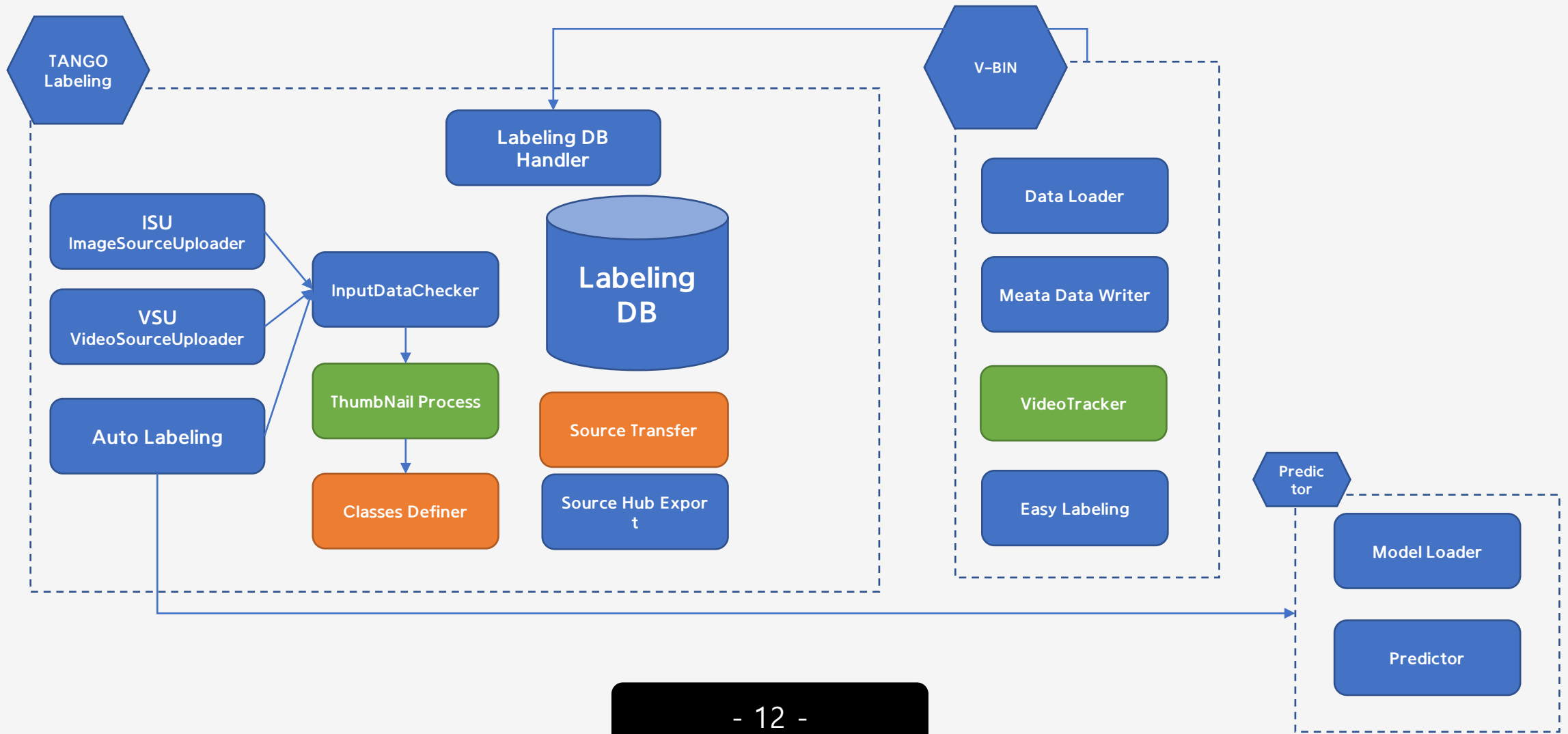
4. TANGO 데이터 레이블링

The screenshot displays the TANGO Data Management interface. On the left is a sidebar with navigation options: Project Management, Target Management, Data Management (highlighted), and Visualization. The main area is titled 'Data Management' and shows a list of datasets under the 'DataSet' tab. Each dataset entry includes a thumbnail, name, status, and a 'Ready' indicator.

Below the dataset list, a 'New Dataset' modal is open. It allows creating a new dataset with a title, data type (Image or Video), and purpose (Classification, Detection, or Segmentation). The modal also shows a table of existing datasets with columns for Label, File Name, Size, Type, and Status.

#	Label	File Name	Size	Type	Status
16	Normal	normal_20200101_0001.jpg	272.2 KB	image/jpeg	✓
17	Virus	VIRUS-3873430-0001.jpg	125.3 KB	image/jpeg	✓
18	Virus	VIRUS-4217562-0002.jpg	85.5 KB	image/jpeg	✓
19	Virus	VIRUS-3871609-0002.jpg	143.1 KB	image/jpeg	✓
20	Virus	VIRUS-4127619-0002.jpg	145.5 KB	image/jpeg	✓
21	Virus	VIRUS-4129462-0002.jpg	188.4 KB	image/jpeg	✓

4. TANGO 데이터 레이블링 - 서비스 구조



4. TANGO 데이터 레이블링 - 서비스 구조



감사합니다.