

Collect Data



Clean Data\*



Train ML/NN  
Model



Deployment



Evaluate on  
unseen data

# DataLoaders

Highest Level

Not much flexibility

- ImageDataLoader
- SegmentationDataLoader
- TextDataLoader
- TabularDataLoader

# DataBlock

Medium-Level

High Flexibility

Building-Blocks

What we will focus on

- ImageBlock
- MaskBlock
- PointBlock
- BboxBlock
- TextBlock
- TabularPandas

# Pipeline

Lowest Level

Highest flexibility

Define Block

Getters

Split the data

DataLoader

Label the Data



# Define Blocks

## Input and Output

ImageBlock

MaskBlock

PointBlock

BboxBlock

TextBlock

TabularPandas

CategoryBlock

MultiCategoryBlock

TransformBlock

## Getters and Number of Inputs

get\_x

get\_y

get\_items

n\_inp

# Split the Data

RandomSplitter

IndexSplitter

GrandparentSplitter

FuncSplitter

FileSplitter

ColSplitter

MaskSplitter

# Label the Data

parent\_label

RegexLabeller

ColReader

get\_y

# DataLoader

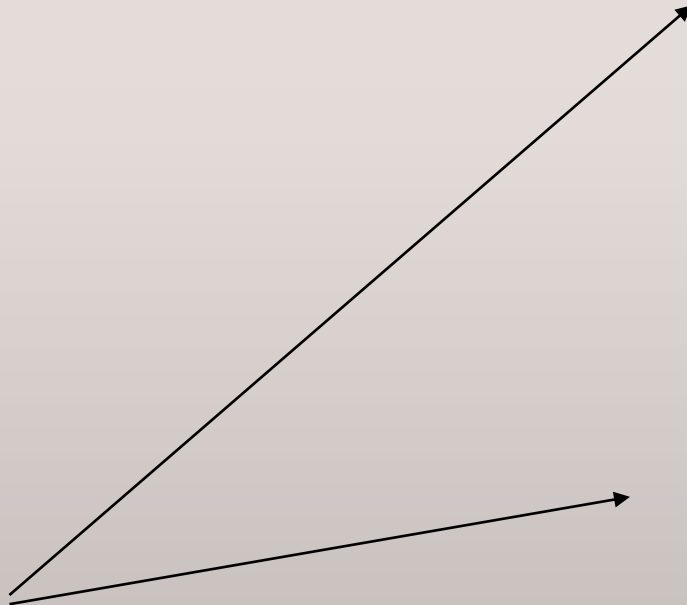
Items/Path

Batch Size

Transforms

Item - CPU

Batch - GPU





# GPU vs CPU Transforms:

- CPU:
  - Performed **first**
  - Need to prepare for a batch (convert to a Tensor, ensure images/text/tabular data can all be collated together, aka they are the same size)
- GPU:
  - Apply to entire batches after collated together, like further resizing, cropping, normalizing the data, and other data augmentation