Flash

PyTorch Lightning

Mar 10, 2021

GET STARTED:

CHAPTER

ONE

QUICK START

Flash is a high-level deep learning framework for fast prototyping, baselining, finetuning and solving deep learning problems. It features a set of tasks for you to use for inference and finetuning out of the box, and an easy to implement API to customize every step of the process for full flexibility.

Flash is built for beginners with a simple API that requires very little deep learning background, and for data scientists, kagglers, applied ML practitioners and deep learning researchers that want a quick way to get a deep learning baseline with advnaced features Pytorch Lightning offers.

1.1 Why Flash?

1.1.1 For getting started with Deep Learning

Easy to learn

If you are just getting started with deep learning, Flash offers common deep learning tasks you can use out-of-the-box in a few lines of code, no math, fancy nn.Modules or research experience required!

Easy to scale

Flash is built on top of Pytorch Lightning, a powerful deep learning research framework for training models at scale. With the power of Lightning, you can train your flash tasks on any hardware: CPUs, GPUs or TPUs without any code changes.

Easy to upskill

If you want create more complex and custmoized models, you can refactor any part of flash with PyTorch or Pytorch Lightning components to get all the flexibility you need. Lightning is just organized PyTorch with the unecessary engineering details abstracted away.

- Flash (high level)
- Lightning (mid-level)
- PyTorch (low-level)

When you need more flexibility you can build your own tasks or simply use Lightning directly.

1.1.2 For Deep learning research

Quickest way to a baseline

Pytorch Lightning is designed to abstract away unnecessary boilerplate, while enabling maximal flexibility. In order to provide full flexibility, solving very common deep learning problems such as classification in Lightning still requires some boilerplate. It can still take quite some time to get a baseline model running on a new dataset or out of domain task. We created Flash to answer our users need for a super quick way to baseline for Lightning using proven backbones for common data patterns. Flash aims to be the easiest starting point for your research- start with a Flash Task to benchmark against, and override any part of flash with Lightning or PyTorch components on your way to SOTA research.

Flexibility where you want it

Flash tasks are essentialy LightningModules, and the Flash Trainer is a thin wrapper for the Lightning Trainer. You can use your own LightningModule instead of the Flash task, the Lightning Trainer instead of the flash trainer, etc. Flash helps you focus even more only on your research, and less on anything else.

Standard best practices

Flash tasks implement the standard best practices for a variety of diffrent models and domains, to save you time digging through different implementations. Flash abstracts even more details than lightning, allowing deep learning experts to share their tips and tricks for solving scoped deep learning problems.

Tip: Read *here* to understand when to use Flash vs Lightning.

1.2 Install

You can install flash using pip or conda:

```
pip install lightning-flash -U
```

1.3 Tasks

Flash is comprised of a collection of Tasks. The Flash tasks are laser-focused objects designed to solve a well-defined type of problem, using state-of-the-art methods.

The Flash tasks contain all the relevant information to solve the task at hand- the number of class labels you want to predict, number of columns in your dataset, as well as details on the model architecture used such as loss function, optimizers, etc.

Here are examples of tasks:

```
from flash.text import TextClassifier
from flash.vision import ImageClassifier
from flash.tabular import TabularClassifier
```

Note: Tasks are inflexible by definition! To get more flexibility, you can simply use `LightningModule` directly or modify and existing task in just a few lines.

1.4 Inference

Inference is the process of generating predictions from trained models. To use a task for inference:

1. Init your task with pretrained weights using a checkpoint (a checkpoint is simply a file that capture the exact value of all parameters used by a model). Local file or URL works.
2. Pass in the data to *flash.core.model.Task.predict()*.

Here's an example of inference.

```
# import our libraries
from flash.text import TextClassifier

# 1. Init the finetuned task from URL
model = TextClassifier.load_from_checkpoint("https://flash-weights.s3.amazonaws.com/
↪text_classification_model.pt")

# 2. Perform inference from list of sequences
predictions = model.predict([
    "Turgid dialogue, feeble characterization - Harvey Keitel a judge?.",
    "The worst movie in the history of cinema.",
    "This guy has done a great job with this movie!",
])

# Expect [0,0, 1] which means [negative, negative, positive]
print(predictions)
```

1.5 Finetune

Finetuning (or transfer-learning) is the process of tweaking a model trained on a large dataset, to your particular (likely much smaller) dataset. To use a Task for finetuning:

1. Download and set up your own data (`DataLoader` or LightningModule work).
2. Init your task.
3. Init a *flash.core.trainer.Trainer* (or a Lightning Trainer).
4. Call *flash.core.trainer.Trainer.finetune()* with your data set.
5. Use your finetuned model for predictions

Here's an example of finetuning.

```
import flash
from flash import download_data
from flash.vision import ImageClassificationData, ImageClassifier

# 1. Download the data
download_data("https://pl-flash-data.s3.amazonaws.com/hymenoptera_data.zip", 'data/')

# 2. Load the data from folders
datamodule = ImageClassificationData.from_folders(
    backbone="resnet18",
    train_folder="data/hymenoptera_data/train/",
    valid_folder="data/hymenoptera_data/val/",
    test_folder="data/hymenoptera_data/test/",
)

# 3. Build the model using desired Task
model = ImageClassifier(num_classes=datamodule.num_classes)

# 4. Create the trainer (run one epoch for demo)
trainer = flash.Trainer(max_epochs=1)

# 5. Finetune the model
trainer.finetune(model, datamodule=datamodule, strategy="freeze")

# 6. Use the model for predictions
predictions = model.predict('data/hymenoptera_data/val/bees/65038344_52a45d090d.jpg')
# Expact 1 -> bee
print(predictions)

predictions = model.predict('data/hymenoptera_data/val/ants/2255445811_dabcdf7258.jpg
↪')
# Expact 0 -> ant
print(predictions)

# 7. Save the new model!
trainer.save_checkpoint("image_classification_model.pt")
```

Once your model is finetuned, use it for prediction anywhere you want!

```
from flash.vision import ImageClassifier

# load finetuned checkpoint
model = ImageClassifier.load_from_checkpoint("image_classification_model.pt")

predictions = model.predict('path/to/your/own/image.png')
```

1.6 Train

When you have enough data, you're likely better off training from scratch instead of finetuning. Steps here are similar to finetune:

1. Download and set up your own data (`DataLoader` or LightningModule work).
2. Init your task.
3. Init a *`flash.core.trainer.Trainer`* (or a Lightning Trainer).
4. Call *`flash.core.trainer.Trainer.fit()`* with your data set.
5. Use your finetuned model for predictions

1.7 A few Built-in Tasks

- *Generic Flash Task*
- *ImageClassification*
- *ImageEmbedder*
- *TextClassification*
- *SummarizationTask*
- *TranslationTask*
- *TabularClassification*

More tasks coming soon!

1.7.1 Contribute a task

The lightning + Flash team is hard at work building more tasks for common deep-learning use cases. But we're looking for incredible contributors like you to submit new tasks!

Join our Slack to get help becoming a contributor!

CHAPTER

TWO

INSTALLATION

Flash is tested on Python 3.6+, and PyTorch 1.6

2.1 Install with pip/conda

```
pip install lightning-flash -U
```

2.2 Install from source

```
pip install git+https://github.com/PyTorchLightning/lightning-flash.git
```

CHAPTER

THREE

TUTORIAL: CREATING A CUSTOM TASK

In this tutorial we will go over the process of creating a custom task, along with a custom data module.

```
import flash

import torch
from torch.utils.data import TensorDataset, DataLoader
from torch import nn
from sklearn import datasets
from sklearn.model_selection import train_test_split
```

3.1 The Task: Linear regression

Here we create a basic linear regression task by subclassing `flash.Task`. For the majority of tasks, you will likely only need to override the `__init__` and `forward` methods.

```
class LinearRegression(flash.Task):
    def __init__(self, num_inputs, learning_rate=0.001, metrics=None):
        # what kind of model do we want?
        model = nn.Linear(num_inputs, 1)

        # what loss function do we want?
        loss_fn = torch.nn.functional.mse_loss

        # what optimizer to do we want?
        optimizer = torch.optim.SGD

        super().__init__(
            model=model,
            loss_fn=loss_fn,
            optimizer=optimizer,
            metrics=metrics,
            learning_rate=learning_rate,
        )

    def forward(self, x):
        # we don't actually need to override this method for this example
        return self.model(x)
```

3.1.1 Where is the training step?

Most models can be trained simply by passing the output of `forward` to the supplied `loss_fn`, and then passing the resulting loss to the supplied `optimizer`. If you need a more custom configuration, you can override `step` (which is called for training, validation, and testing) or override `training_step`, `validation_step`, and `test_step` individually. These methods behave identically to PyTorch Lightning's methods.

3.2 The Data

For a task you will likely need a specific way of loading data. For this example, lets say we want a `flash.DataModule` to be used explicitly for the prediction of diabetes disease progression. We can create this `DataModule` below, wrapping the scikit-learn Diabetes dataset.

```
class DiabetesPipeline(flash.core.data.TaskDataPipeline):
    def after_uncollate(self, samples):
        return [f"disease progression: {float(s):.2f}" for s in samples]

class DiabetesData(flash.DataModule):
    def __init__(self, batch_size=64, num_workers=0):
        x, y = datasets.load_diabetes(return_X_y=True)
        x = torch.from_numpy(x).float()
        y = torch.from_numpy(y).float().unsqueeze(1)
        x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=.20,␣
↪random_state=0)

        train_ds = TensorDataset(x_train, y_train)
        test_ds = TensorDataset(x_test, y_test)

        super().__init__(
            train_ds=train_ds,
            test_ds=test_ds,
            batch_size=batch_size,
            num_workers=num_workers
        )
        self.num_inputs = x.shape[1]

    @staticmethod
    def default_pipeline():
        return DiabetesPipeline()
```

You'll notice we added a `DataPipeline`, which will be used when we call `.predict()` on our model. In this case we want to nicely format our ouput from the model with the string `"disease progression"`, but you could do any sort of post processing you want (see *DataPipeline*).

3.3 Fit

Like any Flash Task, we can fit our model using the `flash.Trainer` by supplying the task itself, and the associated data:

```
data = DiabetesData()
model = LinearRegression(num_inputs=data.num_inputs)

trainer = flash.Trainer(max_epochs=1000)
trainer.fit(model, data)
```

With a trained model we can now perform inference. Here we will use a few examples from the test set of our data:

```
predict_data = torch.tensor([
    [ 0.0199,  0.0507,  0.1048,  0.0701, -0.0360, -0.0267, -0.0250, -0.0026, 0.0037, ␣
↪0.0403],
    [-0.0128, -0.0446,  0.0606,  0.0529,  0.0480,  0.0294, -0.0176,  0.0343, 0.0702, ␣
↪0.0072],
    [ 0.0381,  0.0507,  0.0089,  0.0425, -0.0428, -0.0210, -0.0397, -0.0026, -0.0181,␣
↪ 0.0072],
    [-0.0128, -0.0446, -0.0235, -0.0401, -0.0167,  0.0046, -0.0176, -0.0026, -0.0385,␣
↪-0.0384],
    [-0.0237, -0.0446,  0.0455,  0.0907, -0.0181, -0.0354,  0.0707, -0.0395, -0.0345,␣
↪-0.0094]])

model.predict(predict_data)
```

Because of our custom data pipeline's `after_uncollate` method, we will get a nicely formatted output like the following:

```
['disease progression: 155.90',
 'disease progression: 156.59',
 'disease progression: 152.69',
 'disease progression: 149.05',
 'disease progression: 150.90']
```

CHAPTER

FOUR

FROM FLASH TO LIGHTNING

Flash is built on top of Pytorch Lightning to abstract away the unecessary boilerplate for:

- Data science
- Kaggle
- Business use cases
- Applied research

Flash is a HIGH level library and Lightning is a LOW level library.

- Flash (High-level)
- Lightning (medium-level)
- PyTorch (low-level)

As the complexity increases or decreases, users can move between Flash and Lightning seamlessly to find the level of abstraction that works for them.

Table 1: Abstraction levels

Approach	Flexibility	Minimum DL Expertise level	PyTorch Knowledge	Use cases
Using an out-of-the-box task	Low	Novice+	Low+	Fast baseline, Data Science, Analysis, Applied Research
Using the Generic Task	Medium	Intermediate+	Intermediate+	Fast baseline, data science
Building a custom task	High	Intermediate+	Intermediate+	Fast baseline, custom business context, applied research
Building a LightningModule	Ultimate (organized PyTorch)	Expert+	Expert+	For anything you can do with PyTorch, AI research (academic and corporate)

4.1 Using an out-of-the-box task

Tasks can come from a variety of places:

- Flash
- Other Lightning-based libraries
- Your own library

Using a task requires almost zero knowledge of deep learning and PyTorch. The focus is on solving a problem as quickly as possible. This is great for:

- data science
- analysis
- applied research

4.2 Using the Generic Task

If you encounter a problem that does not have a matching task, you can use the generic task. However, this does require a bit of PyTorch knowledge but not a lot of knowledge over all the details of deep learning.

This is great for:

- data science
- kaggle baselines
- a quick baseline
- applied research
- learning about deep learning

Note: If you've used something like Keras, this is the most similar level of abstraction.

4.3 Building a custom task

If you're feeling adventurous and there isn't an out-of-the-box task for a particular applied problem, consider building your own task. This requires a decent amount of PyTorch knowledge, but not too much because tasks are LightningModules that already abstract a lot of the details for you.

This is great for:

- data science
- researchers building for corporate data science teams
- applied research
- custom business context

Note: In a company setting, a good setup here is to have your own Flash-like library with tasks contextualized with your business problems.

4.4 Building a LightningModule

Once you've reached the threshold of flexibility offered by Flash, it's time to move to a LightningModule directly. LightningModule is organized PyTorch but gives you the same flexibility. However, you must already know PyTorch fairly well and be comfortable with at least basic deep learning concepts.

This is great for:

- experts
- academic AI research
- corporate AI research
- advanced applied research
- publishing papers

CHAPTER

FIVE

GENERAL TASK

A majority of data science problems that involve machine learning can be tackled using Task. With Task you can:

- Pass an arbitrary model
- Pass an arbitrary loss
- Pass an arbitrary optimizer

5.1 Example: Image Classification

```
from flash import Task
from torch import nn, optim
from torch.utils.data import DataLoader, random_split
from torchvision import transforms, datasets
import pytorch_lightning as pl

# model
model = nn.Sequential(
    nn.Flatten(),
    nn.Linear(28 * 28, 128),
    nn.ReLU(),
    nn.Linear(128, 10)
)

# data
dataset = datasets.MNIST('./data_folder', download=True, transform=transforms.
↪ToTensor())
train, val = random_split(dataset, [55000, 5000])

# task
classifier = Task(model, loss_fn=nn.functional.cross_entropy, optimizer=optim.Adam)

# train
pl.Trainer().fit(classifier, DataLoader(train), DataLoader(val))
```

5.2 API reference

5.2.1 Task

class flash.core.**Task**(*model=None*, *loss_fn=None*, *optimizer=torch.optim.Adam*, *metrics=None*, *learning_rate=5e-05*)

A general Task.

Parameters

- **model** (Optional[Module]) – Model to use for the task.
- **loss_fn** (Union[Callable, Mapping, Sequence, None]) – Loss function for training
- **optimizer** (Type[Optimizer]) – Optimizer to use for training, defaults to *torch.optim.SGD*.
- **metrics** (Union[Metric, Mapping, Sequence, None]) – Metrics to compute for training and evaluation.
- **learning_rate** (float) – Learning rate to use for training, defaults to *5e-5*

static default_pipeline()

Pipeline to use when there is no datamodule or it has not defined its pipeline

Return type DataPipeline

predict(*x*, *batch_idx=None*, *skip_collate_fn=False*, *dataloader_idx=None*, *data_pipeline=None*)

Predict function for raw data or processed data

Parameters

- **x** (Any) – Input to predict. Can be raw data or processed data. If str, assumed to be a folder of data.
- **batch_idx** (Optional[int]) – Batch index
- **dataloader_idx** (Optional[int]) – Dataloader index
- **skip_collate_fn** (bool) – Whether to skip the collate step. this is required when passing data already processed for the model, for example, data from a dataloader
- **data_pipeline** (Optional[DataPipeline]) – Use this to override the current data pipeline

Return type Any

Returns The post-processed model predictions

step(*batch*, *batch_idx*)

The training/validation/test step. Override for custom behavior.

Return type Any

CHAPTER

SIX

IMAGE CLASSIFICATION

6.1 The task

The task of identifying what is in an image is called image classification. Typically, Image Classification is used to identify images containing a single object. The task predicts which 'class' the image most likely belongs to with a degree of certainty. A class is a label that desecribes what is in an image, such as 'car', 'house', 'cat' etc. For example, we can train the image classifier task on images of ants and it will learn to predict the probability that an image contains an ant.

6.2 Inference

The *ImageClassifier* is already pre-trained on ImageNet, a dataset of over 14 million images.

Use the *ImageClassifier* pretrained model for inference on any string sequence using `predict()`:

```
# import our libraries
from flash import Trainer
from flash import download_data
from flash.vision import ImageClassificationData, ImageClassifier

# 1. Download the data
download_data("https://pl-flash-data.s3.amazonaws.com/hymenoptera_data.zip", "data/")

# 2. Load the model from a checkpoint
model = ImageClassifier.load_from_checkpoint(
    "https://flash-weights.s3.amazonaws.com/image_classification_model.pt"
)

# 3a. Predict what's on a few images! ants or bees?
predictions = model.predict([
    "data/hymenoptera_data/val/bees/65038344_52a45d090d.jpg",
    "data/hymenoptera_data/val/bees/590318879_68cf112861.jpg",
    "data/hymenoptera_data/val/ants/540543309_ddbb193ee5.jpg",
])
print(predictions)

# 3b. Or generate predictions with a whole folder!
datamodule = ImageClassificationData.from_folder(folder="data/hymenoptera_data/
↪predict/")
predictions = Trainer().predict(model, datamodule=datamodule)
print(predictions)
```

For more advanced inference options, see *Predictions (inference)*.

6.3 Finetuning

Lets say you wanted to develope a model that could determine whether an image contains **ants** or **bees**, using the hymenoptera dataset. Once we download the data using `download_data()`, all we need is the train data and validation data folders to create the *`ImageClassificationData`*.

Note: The dataset contains `train` and `validation` folders, and then each folder contains a **bees** folder, with pictures of bees, and an **ants** folder with images of, you guessed it, ants.

```
hymenoptera_data
├── train
│   ├── ants
│   │   ├── 0013035.jpg
│   │   ├── 1030023514_aad5c608f9.jpg
│   │   ...
│   └── bees
│       ├── 1092977343_cb42b38d62.jpg
│       ├── 1093831624_fb5fbe2308.jpg
│       ...
└── val
    ├── ants
    │   ├── 10308379_1b6c72e180.jpg
    │   ├── 1053149811_f62a3410d3.jpg
    │   ...
    └── bees
        ├── 1032546534_06907fe3b3.jpg
        ├── 10870992_eebeeb3a12.jpg
        ...
```

Now all we need is three lines of code to build to train our task!

```
import flash
from flash import download_data
from flash.vision import ImageClassificationData, ImageClassifier

# 1. Download the data
download_data("https://pl-flash-data.s3.amazonaws.com/hymenoptera_data.zip", "data/")

# 2. Load the data
datamodule = ImageClassificationData.from_folders(
    train_folder="data/hymenoptera_data/train/",
    valid_folder="data/hymenoptera_data/val/",
    test_folder="data/hymenoptera_data/test/",
)

# 3. Build the model
model = ImageClassifier(backbone="resnet18", num_classes=datamodule.num_classes)

# 4. Create the trainer. Run once on data
trainer = flash.Trainer(max_epochs=1)
```

(continues on next page)

(continued from previous page)

```
# 5. Train the model
trainer.finetune(model, datamodule=datamodule, strategy="freeze_unfreeze")

# 6. Test the model
trainer.test()

# 7. Save it!
trainer.save_checkpoint("image_classification_model.pt")
```

6.4 Changing the backbone

By default, we use a ResNet-18 for image classification. You can change the model run by the task by passing in a different backbone.

Note: When changing the backbone, make sure you pass in the same backbone to the Task and the Data object!

```
# 1. organize the data
data = ImageClassificationData.from_folders(
    backbone="resnet34",
    train_folder="data/hymenoptera_data/train/",
    valid_folder="data/hymenoptera_data/val/"
)

# 2. build the task
task = ImageClassifier(num_classes=2, backbone="resnet34")
```

Available backbones:

- resnet18 (default)
- resnet34
- resnet50
- resnet101
- resnet152
- resnext50_32x4d
- resnext101_32x8d
- mobilenet_v2
- vgg11
- vgg13
- vgg16
- vgg19
- densenet121
- densenet169

- densenet161
- swav-imagenet

6.5 API reference

6.5.1 ImageClassifier

class flash.vision.**ImageClassifier**(*num_classes*, *backbone='resnet18'*, *pretrained=True*, *loss_fn=torch.nn.functional.cross_entropy*, *optimizer=torch.optim.SGD*, *metrics=pytorch_lightning.metrics.Accuracy*, *learning_rate=0.001*)

Task that classifies images.

Parameters

- **num_classes** (int) – Number of classes to classify.
- **backbone** (str) – A model to use to compute image features, defaults to "resnet18".
- **pretrained** (bool) – Use a pretrained backbone, defaults to True.
- **loss_fn** (Callable) – Loss function for training, defaults to torch.nn.functional.cross_entropy().
- **optimizer** (Type[Optimizer]) – Optimizer to use for training, defaults to torch.optim.SGD.
- **metrics** (Union[Callable, Mapping, Sequence, None]) – Metrics to compute for training and evaluation, defaults to pytorch_lightning.metrics.Accuracy.
- **learning_rate** (float) – Learning rate to use for training, defaults to 1e-3.

static default_pipeline()

Pipeline to use when there is no datamodule or it has not defined its pipeline

Return type ImageClassificationDataPipeline

6.5.2 ImageClassificationData

class flash.vision.**ImageClassificationData**(*train_ds=None*, *valid_ds=None*, *test_ds=None*, *batch_size=1*, *num_workers=None*)

Data module for image classification tasks.

classmethod ImageClassificationData.**from_filepaths**(*train_filepaths=None*, *train_labels=None*, *train_transform=torchvision.transforms.Compose*, *valid_split=None*, *valid_filepaths=None*, *valid_labels=None*, *valid_transform=torchvision.transforms.Compose*, *test_filepaths=None*, *test_labels=None*, *loader=<function _pil_loader>*, *batch_size=64*, *num_workers=None*, *seed=1234*, ***kwargs*)

Creates a ImageClassificationData object from lists of image filepaths and labels

Parameters

- **train_filepaths** (Union[str, Sequence[Union[str, Path]], None]) – string or sequence of file paths for training dataset. Defaults to None.
- **train_labels** (Optional[Sequence]) – sequence of labels for training dataset. Defaults to None.
- **train_transform** (Optional[Callable]) – transforms for training dataset. Defaults to None.
- **valid_split** (Optional[float]) – if not None, generates val split from train dataloader using this value.
- **valid_filepaths** (Union[str, Sequence[Union[str, Path]], None]) – string or sequence of file paths for validation dataset. Defaults to None.
- **valid_labels** (Optional[Sequence]) – sequence of labels for validation dataset. Defaults to None.
- **valid_transform** (Optional[Callable]) – transforms for validation and testing dataset. Defaults to None.
- **test_filepaths** (Union[str, Sequence[Union[str, Path]], None]) – string or sequence of file paths for test dataset. Defaults to None.
- **test_labels** (Optional[Sequence]) – sequence of labels for test dataset. Defaults to None.
- **loader** (Callable) – function to load an image file. Defaults to None.
- **batch_size** (int) – the batchsize to use for parallel loading. Defaults to 64.
- **num_workers** (Optional[int]) – The number of workers to use for parallelized loading. Defaults to None which equals the number of available CPU threads.
- **seed** (int) – Used for the train/val splits when valid_split is not None

Returns The constructed data module.

Return type *ImageClassificationData*

Examples

```
>>> img_data = ImageClassificationData.from_filepaths(["a.png", "b.png"], [0, 1])
```

Example when labels are in .csv file:

```
train_labels = labels_from_categorical_csv('path/to/train.csv', 'my_id')
valid_labels = labels_from_categorical_csv(path/to/valid.csv', 'my_id')
test_labels = labels_from_categorical_csv(path/to/tests.csv', 'my_id')

data = ImageClassificationData.from_filepaths(
    batch_size=2,
    train_filepaths='path/to/train',
    train_labels=train_labels,
    valid_filepaths='path/to/valid',
    valid_labels=valid_labels,
    test_filepaths='path/to/test',
    test_labels=test_labels,
)
```

classmethod ImageClassificationData.**from_folders**(*train_folder*, *train_transform=torchvision.transforms.Compose*, *valid_folder=None*, *valid_transform=torchvision.transforms.Compose*, *test_folder=None*, *loader=<function _pil_loader>*, *batch_size=4*, *num_workers=None*, ***kwargs*)

Creates a ImageClassificationData object from folders of images arranged in this way:

```
train/dog/xxx.png
train/dog/xxy.png
train/dog/xxz.png
train/cat/123.png
train/cat/nsdf3.png
train/cat/asd932.png
```

Parameters

- **train_folder** (Union[str, Path, None]) – Path to training folder.
- **train_transform** (Optional[Callable]) – Image transform to use for training set.
- **valid_folder** (Union[str, Path, None]) – Path to validation folder.
- **valid_transform** (Optional[Callable]) – Image transform to use for validation and test set.
- **test_folder** (Union[str, Path, None]) – Path to test folder.
- **loader** (Callable) – A function to load an image given its path.
- **batch_size** (int) – Batch size for data loading.
- **num_workers** (Optional[int]) – The number of workers to use for parallelized loading. Defaults to None which equals the number of available CPU threads.

Returns the constructed data module

Return type *ImageClassificationData*

Examples

```
>>> img_data = ImageClassificationData.from_folders("train/")
```

classmethod ImageClassificationData.**from_folder**(*folder*, *transform=torchvision.transforms.Compose*, *loader=<function _pil_loader>*, *batch_size=64*, *num_workers=None*, ***kwargs*)

Creates a ImageClassificationData object from folders of images arranged in this way:

```
folder/dog_xxx.png
folder/dog_xxy.png
folder/dog_xxz.png
folder/cat_123.png
folder/cat_nsdf3.png
folder/cat_asd932_.png
```

Parameters

- **folder** (Union[str, Path]) – Path to the data folder.
- **transform** (Optional[Callable]) – Image transform to apply to the data.
- **loader** (Callable) – A function to load an image given its path.
- **batch_size** (int) – Batch size for data loading.
- **num_workers** (Optional[int]) – The number of workers to use for parallelized loading. Defaults to None which equals the number of available CPU threads.

Returns the constructed data module

Return type *ImageClassificationData*

Examples

```
>>> img_data = ImageClassificationData.from_folder("folder/")
```

CHAPTER

SEVEN

IMAGE EMBEDDER

7.1 The task

Image embedding encodes an image into a vector of image features which can be used for anything like clustering, similarity search or classification.

7.2 Inference

The *ImageEmbedder* is already pre-trained on ImageNet, a dataset of over 14 million images.

Use the *ImageEmbedder* pretrained model for inference on any image tensor or image path using `predict()`:

```
from flash.vision import ImageEmbedder

# Load finetuned task
embedder = ImageEmbedder(backbone="resnet18")

# 2. Perform inference on an image file
embeddings = embedder.predict("path/to/image.png")
print(embeddings)
```

Or on a random image tensor

```
# 2. Perform inference on a random image tensor
import torch
images = torch.rand(32, 3, 224, 224)
embeddings = embedder.predict(images)
print(embeddings)
```

For more advanced inference options, see *Predictions (inference)*.

7.3 Finetuning

To tailor this image embedder to your dataset, finetune first.

```
import flash
from flash import download_data
from flash.vision import ImageClassificationData, ImageEmbedder

# 1. Download the data
download_data("https://pl-flash-data.s3.amazonaws.com/hymenoptera_data.zip", "data/")

# 2. Load the data
datamodule = ImageClassificationData.from_folders(
    train_folder="data/hymenoptera_data/train/",
    valid_folder="data/hymenoptera_data/val/",
    test_folder="data/hymenoptera_data/test/",
)

# 3. Build the model
embedder = ImageEmbedder(backbone="resnet18", embedding_dim=128)

# 4. Create the trainer. Run once on data
trainer = flash.Trainer(max_epochs=1)

# 5. Train the model
trainer.finetune(embedder, datamodule=datamodule, strategy="freeze_unfreeze")

# 6. Test the model
trainer.test()

# 7. Save it!
trainer.save_checkpoint("image_embedder_model.pt")
```

7.4 Changing the backbone

By default, we use the encoder from SwAV pretrained on Imagenet via contrastive learning. You can change the model run by the task by passing in a different backbone.

```
# 1. organize the data
data = ImageClassificationData.from_folders(
    train_folder="data/hymenoptera_data/train/",
    valid_folder="data/hymenoptera_data/val/"
)

# 2. build the task
embedder = ImageEmbedder(backbone="resnet34")
```

Backbones available

Table 1: Backbones

backbone	dataset	training method
resnet18	Imagenet	supervised
resnet34	Imagenet	supervised
resnet50	Imagenet	supervised
resnet101	Imagenet	supervised
resnet152	Imagenet	supervised
swav-imagenet	Imagenet	self-supervised (clustering)

7.5 API reference

7.5.1 ImageEmbedder

class flash.vision.**ImageEmbedder**(*embedding_dim=None*, *backbone='swav-imagenet'*, *pretrained=True*, *loss_fn=torch.nn.functional.cross_entropy*, *optimizer=torch.optim.SGD*, *metrics=pytorch_lightning.metrics.Accuracy*, *learning_rate=0.001*, *pooling_fn=torch.max*)

Task that classifies images.

Parameters

- **embedding_dim** (Optional[int]) – Dimension of the embedded vector. None uses the default from the backbone.
- **backbone** (str) – A model to use to extract image features, defaults to "swav-imagenet".
- **pretrained** (bool) – Use a pretrained backbone, defaults to True.
- **loss_fn** (Callable) – Loss function for training and finetuning, defaults to torch.nn.functional.cross_entropy()
- **optimizer** (Type[Optimizer]) – Optimizer to use for training and finetuning, defaults to torch.optim.SGD.
- **metrics** (Union[Callable, Mapping, Sequence, None]) – Metrics to compute for training and evaluation.
- **learning_rate** (float) – Learning rate to use for training, defaults to 1e-3.
- **pooling_fn** (Callable) – Function used to pool image to generate embeddings, defaults to torch.max().

Example

```
>>> import torch
>>> from flash.vision.embedding import ImageEmbedder
>>> embedder = ImageEmbedder(backbone='resnet18')
>>> image = torch.rand(32, 3, 32, 32)
>>> embeddings = embedder(image)
```

static default_pipeline()

Pipeline to use when there is no datamodule or it has not defined its pipeline

Return type ImageEmbedderDataPipeline

CHAPTER

EIGHT

SUMMARIZATION

8.1 The task

Summarization is the task of summarizing text from a larger document/article into a short sentence/description. For example, taking a web article and describing the topic in a short sentence. This task is a subset of Sequence to Sequence tasks, which requires the model to generate a variable length sequence given an input sequence. In our case the article would be our input sequence, and the short description/sentence would be the output sequence from the model.

8.2 Inference

The *SummarizationTask* is already pre-trained on XSUM, a dataset of online British Broadcasting Corporation articles.

Use the *SummarizationTask* pretrained model for inference on any string sequence using *SummarizationTask* *predict* method:

```
# import our libraries
from flash.text import SummarizationTask

# 1. Load the model from a checkpoint
model = SummarizationTask.load_from_checkpoint("https://flash-weights.s3.amazonaws.
↪com/summarization_model_xsum.pt")

# 2. Perform inference from a sequence
predictions = model.predict([
    """
    Camilla bought a box of mangoes with a Brixton Â£10 note, introduced last year to␣
↪try to keep the money of local
    people within the community.The couple were surrounded by shoppers as they walked␣
↪along Electric Avenue.
    They came to Brixton to see work which has started to revitalise the borough.
    It was Charles' first visit to the area since 1996, when he was accompanied by␣
↪the former
    South African president Nelson Mandela.Greengrocer Derek Chong, who has run a␣
↪stall on Electric Avenue
    for 20 years, said Camilla had been ""nice and pleasant"" when she purchased the␣
↪fruit.
    ""She asked me what was nice, what would I recommend, and I said we've got some␣
↪nice mangoes.
    She asked me were they ripe and I said yes - they're from the Dominican Republic."
↪"
```

(continues on next page)

(continued from previous page)

```
    Mr Chong is one of 170 local retailers who accept the Brixton Pound.
    Customers exchange traditional pound coins for Brixton Pounds and then spend them␣
↪at the market
    or in participating shops.
    During the visit, Prince Charles spent time talking to youth worker Marcus West,␣
↪who works with children
    nearby on an estate off Coldharbour Lane. Mr West said:
    ""He's on the level, really down-to-earth. They were very cheery. The prince is a␣
↪lovely man.""
    He added: ""I told him I was working with young kids and he said, 'Keep up all␣
↪the good work.'""
    Prince Charles also visited the Railway Hotel, at the invitation of his charity␣
↪The Prince's Regeneration Trust.
    The trust hopes to restore and refurbish the building,
    where once Jimi Hendrix and The Clash played, as a new community and business␣
↪centre."
    """
])
print(predictions)
```

Or on a given dataset, use `Trainer` *predict* method:

```
# import our libraries
from flash import Trainer
from flash import download_data
from flash.text import SummarizationData, SummarizationTask

# 1. Download data
download_data("https://pl-flash-data.s3.amazonaws.com/xsum.zip", 'data/')

# 2. Load the model from a checkpoint
model = SummarizationTask.load_from_checkpoint("https://flash-weights.s3.amazonaws.
↪com/summarization_model_xsum.pt")

# 3. Create dataset from file
datamodule = SummarizationData.from_file(
    predict_file="data/xsum/predict.csv",
    input="input",
)

# 4. generate summaries
predictions = Trainer().predict(model, datamodule=datamodule)
print(predictions)
```

For more advanced inference options, see *Predictions (inference)*.

8.3 Finetuning

Say you want to finetune to your own summarization data. We use the XSUM dataset as an example which contains a `train.csv` and `valid.csv`, structured like so:

```
input,target
"The researchers have sequenced the genome of a strain of bacterium that causes the␣
↪virulent infection...","A team of UK scientists hopes to shed light on the␣
↪mysteries of bleeding canker, a disease that is threatening the nation's horse␣
↪chestnut trees."
"Knight was shot in the leg by an unknown gunman at Miami's Shore Club where West was␣
↪holding a pre-MTV Awards...",Hip hop star Kanye West is being sued by Death Row␣
↪Records founder Suge Knight over a shooting at a beach party in August 2005.
...
```

In the above the input column represents the long articles/documents, and the target is the short description used as the target.

All we need is three lines of code to train our model!

```
# import our libraries
import flash
from flash import download_data
from flash.text import SummarizationData, SummarizationTask

# 1. Download data
download_data("https://pl-flash-data.s3.amazonaws.com/xsum.zip", 'data/')

# Organize the data
datamodule = SummarizationData.from_files(
    train_file="data/xsum/train.csv",
    valid_file="data/xsum/valid.csv",
    test_file="data/xsum/test.csv",
    input="input",
    target="target"
)

# 2. Build the task
model = SummarizationTask()

# 4. Create trainer
trainer = flash.Trainer(max_epochs=1, gpus=1)

# 5. Finetune the task
trainer.finetune(model, datamodule=datamodule)

# 6. Save trainer task
trainer.save_checkpoint("summarization_model_xsum.pt")
```

To run the example:

```
python flash_examples/finetuning/summarization.py
```

8.4 Changing the backbone

By default, we use the t5 model for summarization. You can change the model run by the task to any summarization model from HuggingFace/transformers by passing in a backbone parameter.

Note: When changing the backbone, make sure you pass in the same backbone to the Task and the Data object! Since this is a Seq2Seq task, make sure you use a Seq2Seq model.

```
# use google/mt5-small, covering 101 languages
datamodule = SummarizationData.from_files(
    backbone="google/mt5-small",
    train_file="data/wmt_en_ro/train.csv",
    valid_file="data/wmt_en_ro/valid.csv",
    test_file="data/wmt_en_ro/test.csv",
    input="input",
    target="target",
)

model = SummarizationTask(backbone="google/mt5-small")
```

8.5 API reference

8.5.1 SummarizationTask

class flash.text.**SummarizationTask**(*backbone='t5-small'*, *loss_fn=None*, *optimizer=torch.optim.Adam*, *metrics=None*, *learning_rate=5e-05*, *val_target_max_length=None*, *num_beams=4*, *use_stemmer=True*, *rouge_newline_sep=True*)

Task for Seq2Seq Summarization.

Parameters

- **backbone** (str) – backbone model to use for the task.
- **loss_fn** (Union[Callable, Mapping, Sequence, None]) – Loss function for training.
- **optimizer** (Type[Optimizer]) – Optimizer to use for training, defaults to *torch.optim.Adam*.
- **metrics** (Union[Metric, Mapping, Sequence, None]) – Metrics to compute for training and evaluation.
- **learning_rate** (float) – Learning rate to use for training, defaults to *3e-4*
- **val_target_max_length** (Optional[int]) – Maximum length of targets in validation. Defaults to *128*
- **num_beams** (Optional[int]) – Number of beams to use in validation when generating predictions. Defaults to *4*
- **use_stemmer** (bool) – Whether Porter stemmer should be used to strip word suffixes to improve matching.

- **rouge_newline_sep** (bool) – Add a new line at the beginning of each sentence in Rouge Metric calculation.

property task

Override to define AutoConfig task specific parameters stored within the model.

Return type str

8.5.2 SummarizationData

class flash.text.**SummarizationData**(*train_ds=None*, *valid_ds=None*, *test_ds=None*, *batch_size=1*, *num_workers=None*)

classmethod SummarizationData.**from_files**(*train_file*, *input='input'*, *target=None*, *filetype='csv'*, *backbone='t5-small'*, *valid_file=None*, *test_file=None*, *max_source_length=512*, *max_target_length=128*, *padding='max_length'*, *batch_size=16*, *num_workers=None*)

Creates a SummarizationData object from files.

Parameters

- **train_file** (str) – Path to training data.
- **input** (str) – The field storing the source translation text.
- **target** (Optional[str]) – The field storing the target translation text.
- **filetype** (str) – .csv or .json
- **backbone** (str) – tokenizer to use, can use any HuggingFace tokenizer.
- **valid_file** (Optional[str]) – Path to validation data.
- **test_file** (Optional[str]) – Path to test data.
- **max_source_length** (int) – Maximum length of the source text. Any text longer will be truncated.
- **max_target_length** (int) – Maximum length of the target text. Any text longer will be truncated.
- **padding** (Union[str, bool]) – Padding strategy for batches. Default is pad to maximum length.
- **batch_size** (int) – the batchsize to use for parallel loading. Defaults to 16.
- **num_workers** (Optional[int]) – The number of workers to use for parallelized loading. Defaults to None which equals the number of available CPU threads.

Returns The constructed data module.

Return type *SummarizationData*

Examples:

```
train_df = pd.read_csv("train_data.csv")
tab_data = TabularData.from_df(train_df, target="fraud",
                               numerical_input=["account_value"],
                               categorical_input=["account_type"])
```

CHAPTER

NINE

TEXT CLASSIFICATION

9.1 The task

Text classification is the task of assigning a piece of text (word, sentence or document) an appropriate class, or category. The categories depend on the chosen dataset and can range from topics. For example, we can use text classification to understand the sentiment of a given sentence- if it is positive or negative.

9.2 Inference

The *TextClassifier* is already pre-trained on IMDB, a dataset of highly polarized movie reviews, trained for binary classification- to predict if a given review has a positive or negative sentiment.

Use the *TextClassifier* pretrained model for inference on any string sequence using `predict()`:

```
from pytorch_lightning import Trainer

from flash import download_data
from flash.text import TextClassificationData, TextClassifier


# 1. Download the data
download_data("https://pl-flash-data.s3.amazonaws.com/imdb.zip", 'data/')

# 2. Load the model from a checkpoint
model = TextClassifier.load_from_checkpoint("https://flash-weights.s3.amazonaws.com/
↪text_classification_model.pt")

# 2a. Classify a few sentences! How was the movie?
predictions = model.predict([
    "Turgid dialogue, feeble characterization - Harvey Keitel a judge?.",
    "The worst movie in the history of cinema.",
    "I come from Bulgaria where it 's almost impossible to have a tornado."
    "Very, very afraid"
    "This guy has done a great job with this movie!",
])
print(predictions)

# 2b. Or generate predictions from a sheet file!
datamodule = TextClassificationData.from_file(
    predict_file="data/imdb/predict.csv",
```

(continues on next page)

(continued from previous page)

```
    input="review",
)
predictions = Trainer().predict(model, datamodule=datamodule)
print(predictions)
```

For more advanced inference options, see *Predictions (inference)*.

9.3 Finetuning

Say you wanted to create a model that can predict whether a movie review is **positive** or **negative**. We will be using the IMDB dataset, that contains a `train.csv` and `valid.csv`, structured like so:

```
review,sentiment
"Japanese indie film with humor ... ",positive
"Isaac Florentine has made some ...",negative
"After seeing the low-budget ...",negative
"I've seen the original English version ...",positive
"Hunters chase what they think is a man through ...",negative
...
```

All we need is three lines of code to train our model!

```
import flash
from flash.core.data import download_data
from flash.text import TextClassificationData, TextClassifier

# 1. Download the data
download_data("https://pl-flash-data.s3.amazonaws.com/imdb.zip", 'data/')

# 2. Load the data
datamodule = TextClassificationData.from_files(
    train_file="data/imdb/train.csv",
    valid_file="data/imdb/valid.csv",
    test_file="data/imdb/test.csv",
    input="review",
    target="sentiment",
    batch_size=512
)

# 3. Build the task (using the default backbone="bert-base-cased")
model = TextClassifier(num_classes=datamodule.num_classes)

# 4. Create the trainer. Run once on data
trainer = flash.Trainer(max_epochs=1)

# 5. Finetune the task
trainer.finetune(model, datamodule=datamodule, strategy="freeze_unfreeze")

# 6. Test model
trainer.test()

# 7. Save it!
trainer.save_checkpoint("text_classification_model.pt")
```

To run the example:

```
python flash_examples/finetuning/text_classification.py
```

9.4 Changing the backbone

By default, we use the bert-base-uncased model for text classification. You can change the model run by the task to any BERT model from HuggingFace/transformers by passing in a different backbone.

Note: When changing the backbone, make sure you pass in the same backbone to the Task and the Data object!

```
datamodule = TextClassificationData.from_files(
    backbone="bert-base-chinese",
    train_file="data/imdb/train.csv",
    valid_file="data/imdb/valid.csv",
    input="review",
    target="sentiment",
    batch_size=512
)

task = TextClassifier(backbone="bert-base-chinese", num_classes=datamodule.num_
↪classes)
```

9.5 API reference

9.5.1 TextClassifier

class flash.text.classification.model.**TextClassifier**(*num_classes*, *backbone='prajjwal1/bert-tiny'*, *optimizer=torch.optim.Adam*, *metrics=[pytorch_lightning.metrics.classification.Accuracy*, *learning_rate=0.001*)

Task that classifies text.

Parameters

- **num_classes**¶ (int) – Number of classes to classify.
- **backbone**¶ (str) – A model to use to compute text features can be any BERT model from HuggingFace/transformersimage .
- **optimizer**¶ (Type[Optimizer]) – Optimizer to use for training, defaults to *torch.optim.Adam*.
- **metrics**¶ (Union[Callable, Mapping, Sequence, None]) – Metrics to compute for training and evaluation.

- **learning_rate**¶ (float) – Learning rate to use for training, defaults to *1e-3*

static default_pipeline()
Pipeline to use when there is no datamodule or it has not defined its pipeline

Return type ClassificationDataPipeline

step(*batch*, *batch_idx*)
The training/validation/test step. Override for custom behavior.

Return type dict

9.5.2 TextClassificationData

class flash.text.classification.data.**TextClassificationData**(*train_ds=None*, *valid_ds=None*, *test_ds=None*, *batch_size=1*, *num_workers=None*)
Data module for text classification tasks.

classmethod TextClassificationData.**from_files**(*train_file*, *input*, *target*, *filetype='csv'*, *backbone='prajjwal1/bert-tiny'*, *valid_file=None*, *test_file=None*, *max_length=128*, *batch_size=16*, *num_workers=None*)
Creates a TextClassificationData object from files.

Parameters

- **train_file**¶ – Path to training data.
- **input**¶ – The field storing the text to be classified.
- **target**¶ – The field storing the class id of the associated text.
- **filetype**¶ – .csv or .json
- **backbone**¶ – tokenizer to use, can use any HuggingFace tokenizer.
- **valid_file**¶ – Path to validation data.
- **test_file**¶ – Path to test data.
- **batch_size**¶ (int) – the batchsize to use for parallel loading. Defaults to 64.
- **num_workers**¶ (Optional[int]) – The number of workers to use for parallelized loading. Defaults to None which equals the number of available CPU threads.

Returns The constructed data module.

Return type *TextClassificationData*

Examples:

```
train_df = pd.read_csv("train_data.csv")
tab_data = TabularData.from_df(train_df, target="fraud",
                               numerical_input=["account_value"],
                               categorical_input=["account_type"])
```

CHAPTER

TEN

TABULAR CLASSIFICATION

10.1 The task

Tabular classification is the task of assigning a class to samples of structured or relational data. The Flash Tabular Classification task can be used for multi-class classification, or classification of samples in more than two classes. In the following example, the Tabular data is structured into rows and columns, where columns represent properties or features. The task will learn to predict a single target column.

10.2 Finetuning

Say we want to build a model to predict if a passenger survived on the Titanic. We can organize our data in `.csv` files (exportable from Excel, but you can find the kaggle dataset here):

```
PassengerId,Survived,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
1,0,3,"Braund, Mr. Owen Harris",male,22,1,0,A/5 21171,7.25,,S
3,1,3,"Heikkinen, Miss. Laina",female,26,0,0,STON/O2. 3101282,7.925,,S
5,0,3,"Allen, Mr. William Henry",male,35,0,0,373450,8.05,,S
6,0,3,"Moran, Mr. James",male,,0,0,330877,8.4583,,Q
...
```

We can use the Flash Tabular classification task to predict the probability a passenger survived (1 means survived, 0 otherwise), using the feature columns.

We can create *`TabularData`* from csv files using the *`from_csv()`* method. We will pass in:

- **train_csv**- csv file containing the training data converted to a Pandas DataFrame
- **categorical_input**- a list of the names of columns that contain categorical data (strings or integers)
- **numerical_input**- a list of the names of columns that contain numerical continuous data (floats)
- **target**- the name of the column we want to predict

Next, we create the *`TabularClassifier`* task, using the Data module we created.

```
import flash
from flash import download_data
from flash.tabular import TabularClassifier, TabularData
from pytorch_lightning.metrics.classification import Accuracy, Precision, Recall

# 1. Download the data
```

(continues on next page)

(continued from previous page)

```
download_data("https://pl-flash-data.s3.amazonaws.com/titanic.zip", 'data/')

# 2. Load the data
datamodule = TabularData.from_csv(
    "./data/titanic/titanic.csv",
    test_csv="./data/titanic/test.csv",
    categorical_input=["Sex", "Age", "SibSp", "Parch", "Ticket", "Cabin", "Embarked"],
    numerical_input=["Fare"],
    target="Survived",
    val_size=0.25,
    )

# 3. Build the model
model = TabularClassifier.from_data(datamodule, metrics=[Accuracy(), Precision(),␣
↪Recall()])

# 4. Create the trainer. Run 10 times on data
trainer = flash.Trainer(max_epochs=10)

# 5. Train the model
trainer.fit(model, datamodule=datamodule)

# 6. Test model
trainer.test()

# 7. Save it!
trainer.save_checkpoint("tabular_classification_model.pt")

# 8. Predict!
predictions = model.predict("data/titanic/titanic.csv")
print(predictions)
```

10.3 Inference

You can make predcitions on a pretrained model, that has already been trained for the titanic task:

```
from flash.core.data import download_data
from flash.tabular import TabularClassifier

# 1. Download the data
download_data("https://pl-flash-data.s3.amazonaws.com/titanic.zip", 'data/')

# 2. Load the model from a checkpoint
model = TabularClassifier.load_from_checkpoint(
    "https://flash-weights.s3.amazonaws.com/tabnet_classification_model.pt"
)

# 3. Generate predictions from a sheet file! Who would survive?
predictions = model.predict("data/titanic/titanic.csv")
print(predictions)
```

Or you can finetune your own model and use that for prediction:

```
import flash
from flash import download_data
from flash.tabular import TabularClassifier, TabularData

# 1. Load the data
datamodule = TabularData.from_csv(
    "my_data_file.csv",
    test_csv="./data/titanic/test.csv",
    categorical_input=["Sex", "Age", "SibSp", "Parch", "Ticket", "Cabin", "Embarked"],
    numerical_input=["Fare"],
    target="Survived",
    val_size=0.25,
)

# 3. Build the model
model = TabularClassifier.from_data(datamodule, metrics=[Accuracy(), Precision(),␣
↪Recall()])

# 4. Create the trainer
trainer = flash.Trainer()

# 5. Train the model
trainer.fit(model, datamodule=datamodule)

# 6. Test model
trainer.test()

predictions = model.predict("data/titanic/titanic.csv")
print(predictions)
```

10.4 API reference

10.4.1 TabularClassifier

class flash.tabular.**TabularClassifier**(*num_features*, *num_classes*, *embedding_sizes=None*, *loss_fn=torch.nn.functional.cross_entropy*, *optimizer=torch.optim.Adam*, *metrics=None*, *learning_rate=0.001*, ***tabnet_kwargs*)

Task that classifies table rows.

Parameters

- **num_features** (int) – Number of columns in table (not including target column).
- **num_classes** (int) – Number of classes to classify.
- **embedding_sizes** (Optional[List[Tuple]]) – List of (num_classes, emb_dim) to form categorical embeddings.
- **loss_fn** (Callable) – Loss function for training, defaults to cross entropy.
- **optimizer** (Type[Optimizer]) – Optimizer to use for training, defaults to *torch.optim.Adam*.
- **metrics** (Optional[List[Metric]]) – Metrics to compute for training and evaluation.

- **learning_rate** (float) – Learning rate to use for training, defaults to *1e-3*
- ****tabnet_kwargs** – Optional additional arguments for the TabNet model, see pytorch_tabnet.

static default_pipeline()
Pipeline to use when there is no datamodule or it has not defined its pipeline

Return type DataPipeline

predict(*x*, *batch_idx=None*, *skip_collate_fn=False*, *dataloader_idx=None*, *data_pipeline=None*)
Predict function for raw data or processed data

Parameters

- **x** (Any) – Input to predict. Can be raw data or processed data. If str, assumed to be a folder of data.
- **batch_idx** (Optional[int]) – Batch index
- **dataloader_idx** (Optional[int]) – Dataloader index
- **skip_collate_fn** (bool) – Whether to skip the collate step. this is required when passing data already processed for the model, for example, data from a dataloader
- **data_pipeline** (Optional[DataPipeline]) – Use this to override the current data pipeline

Return type Any

Returns The post-processed model predictions

10.4.2 TabularData

class flash.tabular.**TabularData**(*train_df*, *target*, *categorical_input=None*, *numerical_input=None*, *valid_df=None*, *test_df=None*, *batch_size=2*, *num_workers=None*)
Data module for tabular tasks

classmethod TabularData.**from_csv**(*train_csv*, *target*, *categorical_input=None*, *numerical_input=None*, *valid_csv=None*, *test_csv=None*, *batch_size=8*, *num_workers=None*, *val_size=None*, *test_size=None*, ***pandas_kwargs*)
Creates a TextClassificationData object from pandas DataFrames.

Parameters

- **train_csv** (str) – train data csv file.
- **target** (str) – The column containing the class id.
- **categorical_input** (Optional[List]) – The list of categorical columns.
- **numerical_input** (Optional[List]) – The list of numerical columns.
- **valid_csv** (Optional[str]) – validation data csv file.
- **test_csv** (Optional[str]) – test data csv file.
- **batch_size** (int) – the batchsize to use for parallel loading. Defaults to 64.
- **num_workers** (Optional[int]) – The number of workers to use for parallelized loading. Defaults to None which equals the number of available CPU threads.

- **val_size** (Optional[float]) – float between 0 and 1 to create a validation dataset from train dataset
- **test_size** (Optional[float]) – float between 0 and 1 to create a test dataset from train validation

Returns The constructed data module.

Return type *TabularData*

Examples:

```
text_data = TabularData.from_files("train.csv", label_field="class", text_field=
↪"sentence")
```

classmethod TabularData.**from_df**(*train_df*, *target*, *categorical_input=None*, *numerical_input=None*, *valid_df=None*, *test_df=None*, *batch_size=8*, *num_workers=None*, *val_size=None*, *test_size=None*)

Creates a TabularData object from pandas DataFrames.

Parameters

- **train_df** (DataFrame) – train data DataFrame
- **target** (str) – The column containing the class id.
- **categorical_input** (Optional[List]) – The list of categorical columns.
- **numerical_input** (Optional[List]) – The list of numerical columns.
- **valid_df** (Optional[DataFrame]) – validation data DataFrame
- **test_df** (Optional[DataFrame]) – test data DataFrame
- **batch_size** (int) – the batchsize to use for parallel loading. Defaults to 64.
- **num_workers** (Optional[int]) – The number of workers to use for parallelized loading. Defaults to None which equals the number of available CPU threads.
- **val_size** (Optional[float]) – float between 0 and 1 to create a validation dataset from train dataset
- **test_size** (Optional[float]) – float between 0 and 1 to create a test dataset from train validation

Returns The constructed data module.

Return type *TabularData*

Examples:

```
text_data = TextClassificationData.from_files("train.csv", label_field="class",␣
↪text_field="sentence")
```

CHAPTER

ELEVEN

TRANSLATION

11.1 The Task

Translation is the task of translating text from a source language to another, such as English to Romanian. This task is a subset of Sequence to Sequence tasks, which requires the model to generate a variable length sequence given an input sequence. In our case, the task will take an English sequence as input, and output the same sequence in Romanian.

11.2 Inference

The *TranslationTask* is already pre-trained on WMT16 English/Romanian, a dataset of English to Romanian samples, based on the Europarl corpora.

Use the *TranslationTask* pretrained model for inference on any string sequence using *TranslationTask predict* method:

```
# import our libraries
from flash.text import TranslationTask

# 1. Load the model from a checkpoint
model = TranslationTask.load_from_checkpoint("https://flash-weights.s3.amazonaws.com/
↪translation_model_en_ro.pt")

# 2. Perform inference from list of sequences
predictions = model.predict([
    "BBC News went to meet one of the project's first graduates.",
    "A recession has come as quickly as 11 months after the first rate hike and as␣
↪long as 86 months.",
])
print(predictions)
```

Or on a given dataset, use *Trainer predict* method:

```
# import our libraries
from flash import Trainer
from flash import download_data
from flash.text import TranslationData, TranslationTask

# 1. Download data
download_data("https://pl-flash-data.s3.amazonaws.com/wmt_en_ro.zip", 'data/')
```

(continues on next page)

(continued from previous page)

```
# 2. Load the model from a checkpoint
model = TranslationTask.load_from_checkpoint("https://flash-weights.s3.amazonaws.com/
↪translation_model_en_ro.pt")

# 3. Create dataset from file
datamodule = TranslationData.from_file(
    predict_file="data/wmt_en_ro/predict.csv",
    input="input",
)

# 4. generate translations
predictions = Trainer().predict(model, datamodule=datamodule)
print(predictions)
```

For more advanced inference options, see *Predictions (inference)*.

11.3 Finetuning

Say you want to finetune to your own translation data. We use the English/Romanian WMT16 dataset as an example which contains a `train.csv` and `valid.csv`, structured like so:

```
input,target
"Written statements and oral questions (tabling): see Minutes","Declaraţii scrise şi␣
↪întrebări orale (depunere): consultaţi procesul-verbal"
"Closure of sitting","Ridicarea şedinţei"
...
```

In the above the input/target columns represent the English and Romanian translation respectively.

All we need is three lines of code to train our model! By default, we use a mBART backbone for translation which requires a GPU to train.

```
# import our libraries
import flash
from flash import download_data
from flash.text import TranslationData, TranslationTask

# 1. Download data
download_data("https://pl-flash-data.s3.amazonaws.com/wmt_en_ro.zip", 'data/')

# Organize the data
datamodule = TranslationData.from_files(
    train_file="data/wmt_en_ro/train.csv",
    valid_file="data/wmt_en_ro/valid.csv",
    test_file="data/wmt_en_ro/test.csv",
    input="input",
    target="target",
)

# 2. Build the task
model = TranslationTask()

# 4. Create trainer- in this case we need to run on gpus, `precision=16` boosts speed
```

(continues on next page)

(continued from previous page)

```
trainer = flash.Trainer(max_epochs=5, gpus=1, precision=16)

# 5. Finetune the task
trainer.finetune(model, datamodule=datamodule)

# 6. Save model to checkpoint
trainer.save_checkpoint("translation_model_en_ro.pt")
```

To run the example:

```
python flash_examples/finetuning/translation.py
```

11.4 Changing the backbone

You can change the model run by passing in the backbone parameter.

Note: When changing the backbone, make sure you pass in the same backbone to the Task and the Data object! Since this is a Seq2Seq task, make sure you use a Seq2Seq model.

```
datamodule = TranslationData.from_files(
    backbone="t5-small",
    train_file="data/wmt_en_ro/train.csv",
    valid_file="data/wmt_en_ro/valid.csv",
    test_file="data/wmt_en_ro/test.csv",
    input="input",
    target="target",
)

model = TranslationTask(backbone="t5-small")
```

11.5 API reference

11.5.1 TranslationTask

class flash.text.**TranslationTask**(*backbone='facebook/mbart-large-en-ro'*, *loss_fn=None*, *optimizer=torch.optim.Adam*, *metrics=None*, *learning_rate=0.0003*, *val_target_max_length=128*, *num_beams=4*, *n_gram=4*, *smooth=False*)

Task for Sequence2Sequence Translation.

Parameters

- **backbone** (str) – backbone model to use for the task.
- **loss_fn** (Union[Callable, Mapping, Sequence, None]) – Loss function for training.

- **optimizer**¶ (Type[Optimizer]) – Optimizer to use for training, defaults to *torch.optim.Adam*.
- **metrics**¶ (Union[Metric, Mapping, Sequence, None]) – Metrics to compute for training and evaluation.
- **learning_rate**¶ (float) – Learning rate to use for training, defaults to *3e-4*
- **val_target_max_length**¶ (Optional[int]) – Maximum length of targets in validation. Defaults to *128*
- **num_beams**¶ (Optional[int]) – Number of beams to use in validation when generating predictions. Defaults to *4*
- **n_gram**¶ (bool) – Maximum n_grams to use in metric calculation. Defaults to *4*
- **smooth**¶ (bool) – Apply smoothing in BLEU calculation. Defaults to *True*

property task

Override to define AutoConfig task specific parameters stored within the model.

Return type str

11.5.2 TranslationData

class flash.text.**TranslationData**(*train_ds=None*, *valid_ds=None*, *test_ds=None*, *batch_size=1*, *num_workers=None*)

Data module for Translation tasks.

classmethod TranslationData.**from_files**(*train_file*, *input='input'*, *target=None*, *filetype='csv'*, *backbone='facebook/mbart-large-en-ro'*, *valid_file=None*, *test_file=None*, *max_source_length=128*, *max_target_length=128*, *padding='max_length'*, *batch_size=8*, *num_workers=None*)

Creates a TranslateData object from files.

Parameters

- **train_file**¶ – Path to training data.
- **input**¶ (str) – The field storing the source translation text.
- **target**¶ (Optional[str]) – The field storing the target translation text.
- **filetype**¶ – .csv or .json
- **backbone**¶ – tokenizer to use, can use any HuggingFace tokenizer.
- **valid_file**¶ – Path to validation data.
- **test_file**¶ – Path to test data.
- **max_source_length**¶ (int) – Maximum length of the source text. Any text longer will be truncated.
- **max_target_length**¶ (int) – Maximum length of the target text. Any text longer will be truncated.
- **padding**¶ (Union[str, bool]) – Padding strategy for batches. Default is pad to maximum length.
- **batch_size**¶ (int) – the batchsize to use for parallel loading. Defaults to 8.

- **num_workers** (Optional[int]) – The number of workers to use for parallelized loading. Defaults to None which equals the number of available CPU threads.

Returns The constructed data module.

Return type TranslateData

Examples:

```
train_df = pd.read_csv("train_data.csv")
tab_data = TabularData.from_df(train_df, target="fraud",
                               numerical_input=["account_value"],
                               categorical_input=["account_type"])
```

CHAPTER

TWELVE

OBJECT DETECTION

12.1 The task

The object detection task identifies instances of objects of a certain class within an image.

12.2 Inference

The *ObjectDetector* is already pre-trained on COCO train2017, a dataset with 91 classes (123,287 images, 886,284 instances).

```
annotation{
    "id": int,
    "image_id": int,
    "category_id": int,
    "segmentation": RLE or [polygon],
    "area": float,
    "bbox": [x,y,width,height],
    "iscrowd": 0 or 1,
}

categories[{
    "id": int,
    "name": str,
    "supercategory": str,
}]
```

Use the *ObjectDetector* pretrained model for inference on any image tensor or image path using `predict()`:

```
from flash.vision import ObjectDetector

# 1. Load the model
detector = ObjectDetector()

# 2. Perform inference on an image file
predictions = detector.predict("path/to/image.png")
print(predictions)
```

Or on a random image tensor

```
# Perform inference on a random image tensor
import torch
images = torch.rand(32, 3, 1080, 1920)
predictions = detector.predict(images)
print(predictions)
```

For more advanced inference options, see *Predictions (inference)*.

12.3 Finetuning

To tailor the object detector to your dataset, you would need to have it in COCO Format, and then finetune the model.

Tip: You could also pass *trainable_backbone_layers* to `ObjectDetector` and train the model.

```
import flash
from flash.core.data import download_data
from flash.vision import ObjectDetectionData, ObjectDetector

# 1. Download the data
# Dataset Credit: https://www.kaggle.com/ultralytics/coco128
download_data("https://github.com/zhiqwang/yolov5-rt-stack/releases/download/v0.3.0/
↪coco128.zip", "data/")

# 2. Load the Data
datamodule = ObjectDetectionData.from_coco(
    train_folder="data/coco128/images/train2017/",
    train_ann_file="data/coco128/annotations/instances_train2017.json",
    batch_size=2
)

# 3. Build the model
model = ObjectDetector(model="fasterrcnn", backbone="simclr-imagenet", num_
↪classes=datamodule.num_classes)

# 4. Create the trainer. Run thrice on data
trainer = flash.Trainer(max_epochs=3)

# 5. Finetune the model
trainer.finetune(model, datamodule)

# 6. Save it!
trainer.save_checkpoint("object_detection_model.pt")
```

12.4 Model

By default, we use the Faster R-CNN model with a ResNet-50 FPN backbone. We also support RetinaNet. The inputs could be images of different sizes. The model behaves differently for training and evaluation. For training, it expects both the input tensors as well as the targets. And during the evaluation, it expects only the input tensors and returns predictions for each image. The predictions are a list of boxes, labels, and scores.

12.5 Changing the backbone

By default, we use a ResNet-50 FPN backbone. You can change the backbone for the model by passing in a different backbone.

```
# 1. Organize the data
datamodule = ObjectDetectionData.from_coco(
    train_folder="data/coco128/images/train2017/",
    train_ann_file="data/coco128/annotations/instances_train2017.json",
    batch_size=2
)

# 2. Build the Task
model = ObjectDetector(model="retinanet", backbone="resnet101", num_
↪classes=datamodule.num_classes)
```

Available backbones:

- resnet18
- resnet34
- resnet50
- resnet101
- resnet152
- resnext50_32x4d
- resnext101_32x8d
- mobilenet_v2
- vgg11
- vgg13
- vgg16
- vgg19
- densenet121
- densenet169
- densenet161
- swav-imagenet
- simclr-imagenet

12.6 API reference

12.6.1 ObjectDetector

class flash.vision.**ObjectDetector**(*num_classes*, *model='fasterrcnn'*, *backbone=None*, *fpn=True*, *pretrained=True*, *pretrained_backbone=True*, *trainable_backbone_layers=3*, *anchor_generator=None*, *loss=None*, *metrics=None*, *optimizer=torch.optim.Adam*, *learning_rate=0.001*, ***kwargs*)

Object detection task

Ref: Lightning Bolts https://github.com/PyTorchLightning/pytorch-lightning-bolts

Parameters

- **num_classes**¶ (int) – the number of classes for detection, including background
- **model**¶ (str) – a string of :attr`_models`. Defaults to 'fasterrcnn'.
- **backbone**¶ (Optional[str]) – Pretained backbone CNN architecture. Constructs a model with a ResNet-50-FPN backbone when no backbone is specified.
- **fpn**¶ (bool) – If True, creates a Feature Pyramind Network on top of Resnet based CNNs.
- **pretrained**¶ (bool) – if true, returns a model pre-trained on COCO train2017
- **pretrained_backbone**¶ (bool) – if true, returns a model with backbone pre-trained on Imagenet
- **trainable_backbone_layers**¶ (int) – number of trainable resnet layers starting from final block. Only applicable for *fasterrcnn*.
- **loss**¶ – the function(s) to update the model with. Has no effect for torchvision detection models.
- **metrics**¶ (Union[Callable, Module, Mapping, Sequence, None]) – The provided metrics. All metrics here will be logged to progress bar and the respective logger.
- **optimizer**¶ (Type[Optimizer]) – The optimizer to use for training. Can either be the actual class or the class name.
- **pretrained**¶ – Whether the model from torchvision should be loaded with it's pretrained weights. Has no effect for custom models.
- **learning_rate**¶ (float) – The learning rate to use for training

static default_pipeline()

Pipeline to use when there is no datamodule or it has not defined its pipeline

Return type ObjectDetectionDataPipeline

training_step(*batch*, *batch_idx*)

The training step. Overrides Task.training_step

Return type Any

12.6.2 ObjectDetectionData

class flash.vision.**ObjectDetectionData**(*train_ds=None*, *valid_ds=None*, *test_ds=None*, *batch_size=1*, *num_workers=None*)

classmethod ObjectDetectionData.**from_coco**(*train_folder=None*, *train_ann_file=None*, *train_transform=torchvision.transforms.ToTensor*, *valid_folder=None*, *valid_ann_file=None*, *valid_transform=torchvision.transforms.ToTensor*, *test_folder=None*, *test_ann_file=None*, *test_transform=torchvision.transforms.ToTensor*, *batch_size=4*, *num_workers=None*, ***kwargs*)

CHAPTER

THIRTEEN

MODEL

class flash.core.model.**Task**(*model=None*, *loss_fn=None*, *optimizer=torch.optim.Adam*, *metrics=None*, *learning_rate=5e-05*)

A general Task.

Parameters

- **model** (Optional[Module]) – Model to use for the task.
- **loss_fn** (Union[Callable, Mapping, Sequence, None]) – Loss function for training
- **optimizer** (Type[Optimizer]) – Optimizer to use for training, defaults to *torch.optim.SGD*.
- **metrics** (Union[Metric, Mapping, Sequence, None]) – Metrics to compute for training and evaluation.
- **learning_rate** (float) – Learning rate to use for training, defaults to *5e-5*

static default_pipeline()

Pipeline to use when there is no datamodule or it has not defined its pipeline

Return type DataPipeline

predict(*x*, *batch_idx=None*, *skip_collate_fn=False*, *dataloader_idx=None*, *data_pipeline=None*)

Predict function for raw data or processed data

Parameters

- **x** (Any) – Input to predict. Can be raw data or processed data. If str, assumed to be a folder of data.
- **batch_idx** (Optional[int]) – Batch index
- **dataloader_idx** (Optional[int]) – Dataloader index
- **skip_collate_fn** (bool) – Whether to skip the collate step. this is required when passing data already processed for the model, for example, data from a dataloader
- **data_pipeline** (Optional[DataPipeline]) – Use this to override the current data pipeline

Return type Any

Returns The post-processed model predictions

step(*batch*, *batch_idx*)

The training/validation/test step. Override for custom behavior.

Return type Any

CHAPTER

FOURTEEN

DATA

14.1 DataPipeline

To make tasks work for inference, one must create a `DataPipeline`. The `flash.core.data.DataPipeline` exposes 6 hooks to override:

```
class DataPipeline:
    """
    This class purpose is to facilitate the conversion of raw data to processed or␣
↪batched data and back.
    Several hooks are provided for maximum flexibility.

    collate_fn:
        - before_collate
        - collate
        - after_collate

    uncollate_fn:
        - before_uncollate
        - uncollate
        - after_uncollate
    """

    def before_collate(self, samples: Any) -> Any:
        """Override to apply transformations to samples"""
        return samples

    def collate(self, samples: Any) -> Any:
        """Override to convert a set of samples to a batch"""
        if not isinstance(samples, Tensor):
            return default_collate(samples)
        return samples

    def after_collate(self, batch: Any) -> Any:
        """Override to apply transformations to the batch"""
        return batch

    def before_uncollate(self, batch: Any) -> Any:
        """Override to apply transformations to the batch"""
        return batch

    def uncollate(self, batch: Any) -> ny:
        """Override to convert a batch to a set of samples"""
        samples = batch
```

(continues on next page)

(continued from previous page)

```
        return samples

    def after_uncollate(self, samples: Any) -> Any:
        """Override to apply transformations to samples"""
        return samplesA
```

Use these utilities to download data.

14.2 download_data

flash.core.data.utils.**download_data**(*url*, *path='data/'*)

Downloads data automatically from the given url to the path. Defaults to data/ for the path. Automatically handles .csv, .zip

Example:

```
from flash import download_data
```

Parameters

- **url** (str) – path
- **path** (str) – local

Return type None

CHAPTER

FIFTEEN

TRAINING FROM SCRATCH

Some Flash tasks have been pretrained on large data sets. To accelerate your training, calling the *finetune()* method using a pretrained backbone will fine-tune the backbone to generate a model customized to your data set and desired task. If you want to train the task from scratch instead, pass *pretrained=False* parameter when creating your task. Then, use the *fit()* method to train your model.

```
import flash
from flash import download_data
from flash.vision import ImageClassificationData, ImageClassifier

# 1. download and organize the data
download_data("https://download.pytorch.org/tutorial/hymenoptera_data.zip", 'data/')

data = ImageClassificationData.from_folders(
    train_folder="data/hymenoptera_data/train/",
    valid_folder="data/hymenoptera_data/val/"
)

# 2. build the task, and turn off pre-training
task = ImageClassifier(num_classes=2, pretrained=False)

# 3. train!
trainer = flash.Trainer()
trainer.fit(model, data)
```

15.1 Training options

Flash tasks supports many advanced training functionalities out-of-the-box, such as:

- limit number of epochs

```
# train for 10 epochs
flash.Trainer(max_epochs=10)
```

- Training on GPUs

```
# train on 1 GPU
flash.Trainer(gpus=1)
```

- Training on multiple GPUs

```
# train on multiple GPUs
flash.Trainer(gpus=4)
```

```
# train on gpu 1, 3, 5 (3 gpus total)
flash.Trainer(gpus=[1, 3, 5])
```

- Using mixed precision training

```
# Multi GPU with mixed precision
flash.Trainer(gpus=2, precision=16)
```

- Training on TPUs

```
# Train on TPUs
flash.Trainer(tpu_cores=8)
```

You can add to the flash Trainer any argument from the Lightning trainer! Learn more about the Lightning Trainer here.

15.1.1 Trainer API

class flash.core.trainer.**Trainer**(**args*, ***kwargs*)

finetune(*model*, *train_dataloader=None*, *val_dataloaders=None*, *datamodule=None*, *strategy=None*)

Runs the full optimization routine. Same as pytorch_lightning.Trainer().fit(), but unfreezes layers of the backbone throughout training layers of the backbone throughout training.

Parameters

- **datamodule** (Optional[LightningDataModule]) – A instance of LightningDataModule.
- **model** (LightningModule) – Model to fit.
- **train_dataloader** (Optional[DataLoader]) – A Pytorch DataLoader with training samples. If the model has a predefined train_dataloader method this will be skipped.
- **val_dataloaders** (Union[DataLoader, List[DataLoader], None]) – Either a single Pytorch Dataloader or a list of them, specifying validation samples. If the model has a predefined val_dataloaders method this will be skipped
- **strategy** (Union[str, BaseFinetuning, None]) – Should either be a string or a finetuning callback subclassing pytorch_lightning.callbacks.BaseFinetuning.

  Currently, default strategies can be enabled with these strings:

  - no_freeze,
  - freeze,
  - freeze_unfreeze,
  - unfreeze_milestones

fit(*model*, *train_dataloader=None*, *val_dataloaders=None*, *datamodule=None*)

Runs the full optimization routine. Same as pytorch_lightning.Trainer().fit()

Parameters

- **datamodule¶** (Optional[LightningDataModule]) – A instance of LightningDataModule.
- **model¶** (LightningModule) – Model to fit.
- **train_dataloader¶** (Optional[DataLoader]) – A Pytorch DataLoader with training samples. If the model has a predefined train_dataloader method this will be skipped.
- **val_dataloaders¶** (Union[DataLoader, List[DataLoader], None]) – Either a single Pytorch Dataloader or a list of them, specifying validation samples. If the model has a predefined val_dataloaders method this will be skipped

CHAPTER

SIXTEEN

FINETUNING

Finetuning (or transfer-learning) is the process of tweaking a model trained on a large dataset, to your particular (likely much smaller) dataset.

16.1 Terminology

Here are common terms you need to be familiar with:

Table 1: Terminology

Term	Definition
Finetuning	The process of tweaking a model trained on a large dataset, to your particular (likely much smaller) dataset
Transfer learning	The common name for finetuning
Backbone	The neural network that was pretrained on a different dataset
Head	Another neural network (usually smaller) that maps the backbone to your particular dataset
Freeze	Disabling gradient updates to a model (ie: not learning)
Unfreeze	Enabling gradient updates to a model

16.2 3 steps to finetune in Flash

All Flash tasks have a pre-trained backbone that was already trained on large datasets such as ImageNet. Finetuning on already pretrained models decrease training time significantly.

To finetune using Flash, follow these 3 steps:

1. Load your data and organize it using a DataModule customized for the task.
2. Pick a Task which has all the state-of-the-art built in (example: *`ImageClassifier`*).
3. Choose a Finetune strategy and call the *`finetune()`* method

Here are the steps in code

```
import flash
from flash import download_data
from flash.vision import ImageClassificationData, ImageClassifier

# 1. download and organize the data
download_data("https://download.pytorch.org/tutorial/hymenoptera_data.zip", 'data/')

data = ImageClassificationData.from_folders(
    train_folder="data/hymenoptera_data/train/",
    valid_folder="data/hymenoptera_data/val/"
)

# 2. build the model
model = ImageClassifier(num_classes=2)

# 3. Build the trainer and finetune! In this case, using the no_freeze strategy
trainer = flash.Trainer()
trainer.finetune(task, data, strategy="no_freeze")
```

Tip: If you have a large dataset and prefer to train from scratch, see the *Training from scratch* guide.

16.3 Using a finetuned model

Once you've finetuned, use the model to predict.

```
predictions = task.predict('data/hymenoptera_data/val/bees/65038344_52a45d090d.jpg')
print(predictions)
```

Or use a different checkpoint for prediction

```
# Save the checkpoint while training.
trainer.save_checkpoint("image_classification_model.pt")

# load the finetuned model
classifier = ImageClassifier.load_from_checkpoint('image_classification_model.pt')

# predict!
predictions = classifier.predict('data/hymenoptera_data/val/bees/65038344_52a45d090d.
↪jpg')
print(predictions)
```

16.4 Finetune strategies

Finetuning is very task specific. Each task encodes the best finetuning practices for that task. However, Flash gives you a few default strategies for finetuning.

Finetuning operates on two things, the model backbone and the head. The backbone is the neural network that was pre-trained. The head is another neural network that bridges between the backbone and your particular dataset.

16.4.1 no_freeze

In this strategy, the backbone and the head are unfrozen from the beginning.

```
trainer.finetune(task, data, strategy='no_freeze')
```

In pseudocode, this looks like:

```
backbone = Resnet50()
head = nn.Linear(...)

backbone.unfreeze()
head.unfreeze()

train(backbone, head)
```

16.4.2 freeze

The freeze strategy keeps the backbone frozen throughout.

```
trainer.finetune(task, data, strategy='freeze')
```

The psedocode looks like:

```
backbone = Resnet50()
head = nn.Linear(...)

# freeze backbone
backbone.freeze()
head.unfreeze()

train(backbone, head)
```

16.4.3 freeze_unfreeze

In this strategy, the backbone is frozen for 10 epochs then unfrozen.

```
trainer.finetune(model, data, strategy='freeze_unfreeze')
```

```
from flash.core.finetuning import FreezeUnfreeze

# finetune for 10 epochs. Backbone will be frozen for 5 epochs.
trainer = flash.Trainer(max_epochs=10)
trainer.finetune(model, data, strategy=FreezeUnfreeze(unfreeze_epoch=5))
```

Under the hood, the pseudocode looks like:

```
backbone = Resnet50()
head = nn.Linear(...)

# freeze backbone
backbone.freeze()
head.unfreeze()

train(backbone, head, epochs=10)

# unfreeze after 10 epochs
backbone.unfreeze()

train(backbone, head)
```

16.5 Advanced strategies

Every finetune strategy can also be customized.

16.5.1 freeze_unfreeze

In this strategy, the backbone is frozen for x epochs then unfrozen.

Here we unfreeze the backbone at epoch 11.

```
from flash.core.finetuning import FreezeUnfreeze

trainer = flash.Trainer(max_epochs=10)
trainer.finetune(model, data, strategy=FreezeUnfreeze(unfreeze_epoch=11))
```

16.5.2 unfreeze_milestones

This strategy allows you to unfreeze part of the backbone at predetermined intervals

Here's an example where: - backbone starts frozen - at epoch 3 the last 2 layers unfreeze - at epoch 8 the full backbone unfreezes

```
from flash.core.finetuning import UnfreezeMilestones

# finetune for 10 epochs.
trainer = flash.Trainer(max_epochs=10)
trainer.finetune(model, data, strategy=UnfreezeMilestones(unfreeze_milestones=(3, 8),␣
↪num_layers=2))
```

Under the hood, the pseudocode looks like:

```
backbone = Resnet50()
head = nn.Linear(...)

# freeze backbone
backbone.freeze()
head.unfreeze()

train(backbone, head, epochs=3)

# unfreeze last 2 layers at epoch 3
backbone.unfreeze_last_layers(2)

train(backbone, head, epochs=8)

# unfreeze the full backbone
backbone.unfreeze()
```

16.6 Custom Strategy

For even more customization, create your own finetuning callback. Learn more about callbacks here.

```
from flash.core.finetuning import FlashBaseFinetuning

# Create a finetuning callback
class FeatureExtractorFreezeUnfreeze(FlashBaseFinetuning):

    def __init__(self, unfreeze_at_epoch: int = 5, train_bn: bool = True):
        # this will set self.attr_names as ["feature_extractor"]
        super().__init__("feature_extractor", train_bn)
        self._unfreeze_at_epoch = unfreeze_at_epoch

    def finetune_function(self, pl_module, current_epoch, optimizer, opt_idx):
        # unfreeze any module you want by overriding this function

        # When ``current_epoch`` is 5, feature_extractor will start to be trained.
        if current_epoch == self._unfreeze_at_epoch:
            self.unfreeze_and_add_param_group(
                module=pl_module.feature_extractor,
                optimizer=optimizer,
                train_bn=True,
            )

# Init the trainer
trainer = flash.Trainer(max_epochs=10)

# pass the callback to trainer.finetune
trainer.finetune(model, data, strategy=FeatureExtractorFreezeUnfreeze(unfreeze_
↪epoch=5))
```

CHAPTER

SEVENTEEN

PREDICTIONS (INFERENCE)

You can use Flash to get predictions on pretrained or finetuned models.

17.1 Predict on a single sample of data

You can pass in a sample of data (image file path, a string of text, etc) to the *predict()* method.

```
from flash import Trainer
from flash.core.data import download_data
from flash.vision import ImageClassificationData, ImageClassifier


# 1. Download the data set
download_data("https://pl-flash-data.s3.amazonaws.com/hymenoptera_data.zip", 'data/')

# 2. Load the model from a checkpoint
model = ImageClassifier.load_from_checkpoint("https://flash-weights.s3.amazonaws.com/
↪image_classification_model.pt")

# 3. Predict whether the image contains an ant or a bee
predictions = model.predict("data/hymenoptera_data/val/bees/65038344_52a45d090d.jpg")
print(predictions)
```

17.2 Predict on a csv file

```
from flash.core.data import download_data
from flash.tabular import TabularClassifier

# 1. Download the data
download_data("https://pl-flash-data.s3.amazonaws.com/titanic.zip", 'data/')

# 2. Load the model from a checkpoint
model = TabularClassifier.load_from_checkpoint(
    "https://flash-weights.s3.amazonaws.com/tabnet_classification_model.pt"
)

# 3. Generate predictions from a csv file! Who would survive?
predictions = model.predict("data/titanic/titanic.csv")
print(predictions)
```

CHAPTER

EIGHTEEN

INDICES AND TABLES

- genindex
- modindex
- search

INDEX