

# Add a new dataset

DIGITS

lukash (Logout) Info About

## Home

2/2 GPUs available

No Jobs Running

Datasets (1)

Models (4)

Pretrained Models (1)

Group Jobs: ☒

New Dataset

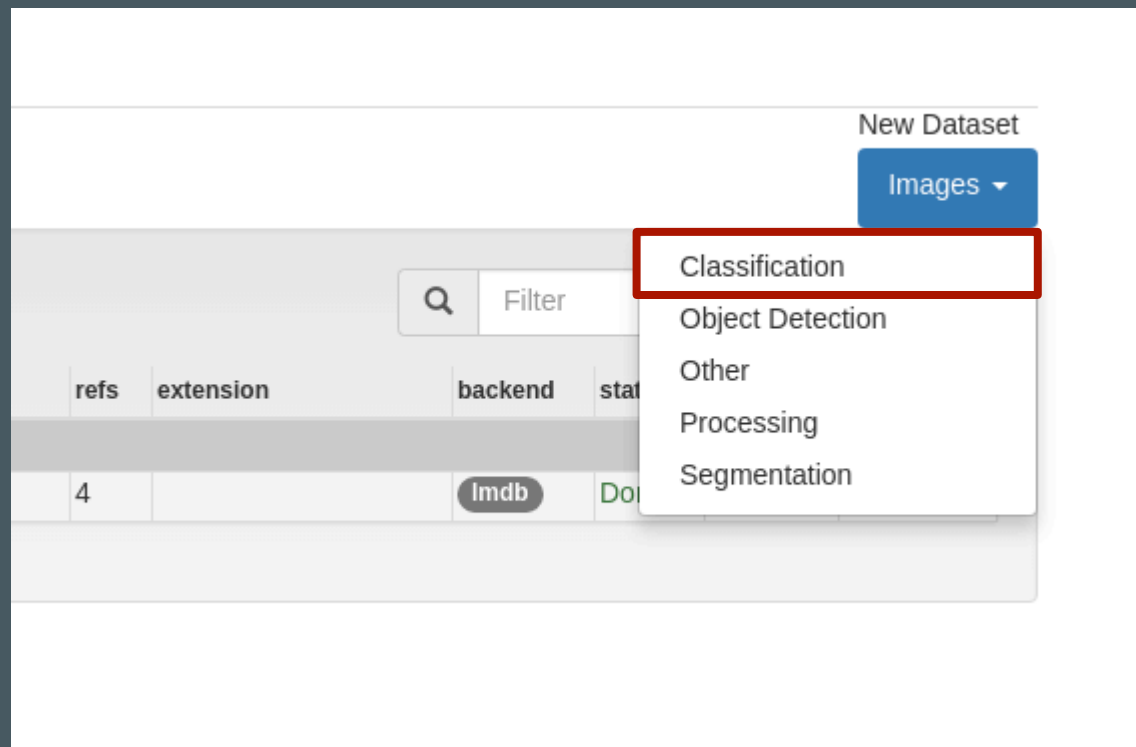
Images

Delete Group

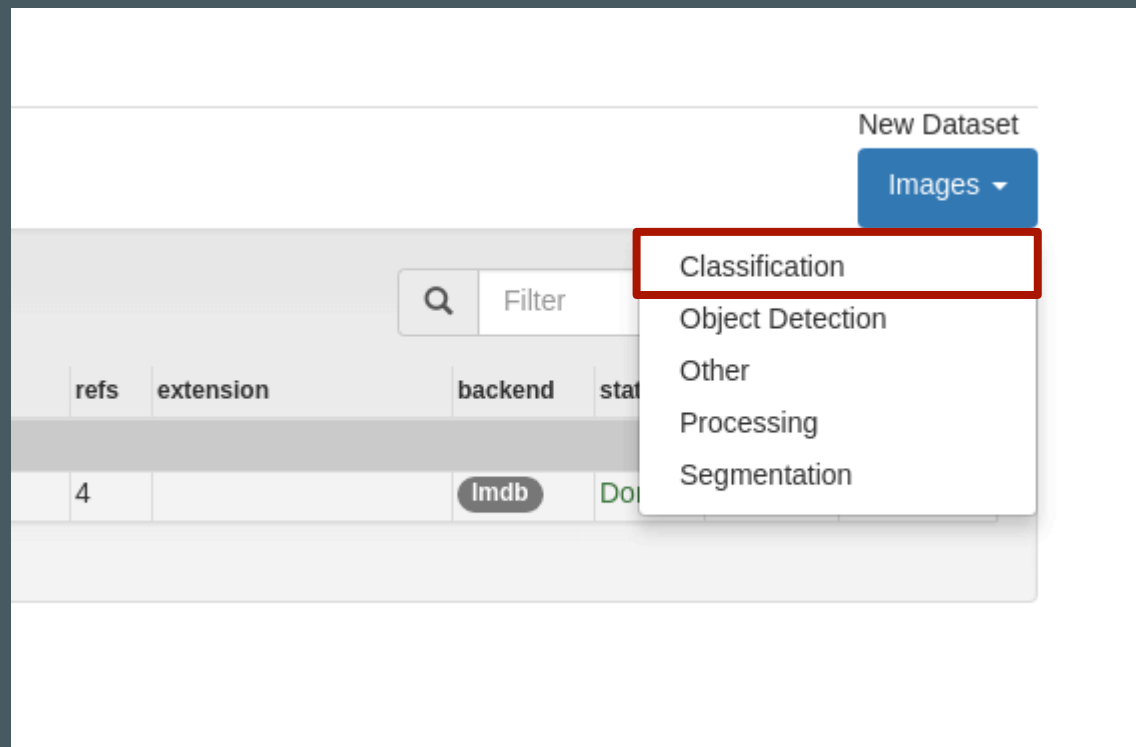
Filter

name	refs	extension	backend	status	elapsed	submitted
------	------	-----------	---------	--------	---------	-----------

# Add a new dataset



# Add a new dataset



# Add a new dataset

**Image Type** ?

Grayscale ▼

**Image size (Width x Height)** ?

28 x 28

**Resize Transformation** ?

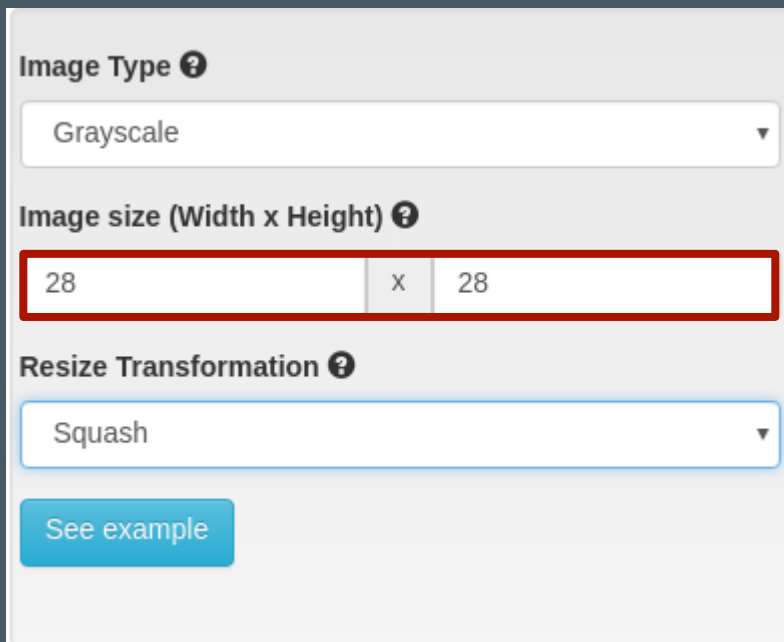
Squash ▼

[See example](#)

Image Type:

- Grayscale : 1 channel
- Color : 3 channels (RGB)

# Add a new dataset



The screenshot shows a configuration panel for adding a new dataset. It has three main sections: 'Image Type' with a dropdown menu set to 'Grayscale'; 'Image size (Width x Height)' with two input fields containing '28' and '28' separated by an 'x' symbol, which is highlighted with a red border; and 'Resize Transformation' with a dropdown menu set to 'Squash'. At the bottom is a blue button labeled 'See example'.

Image Type ?

Grayscale ▼

Image size (Width x Height) ?

28 x 28

Resize Transformation ?

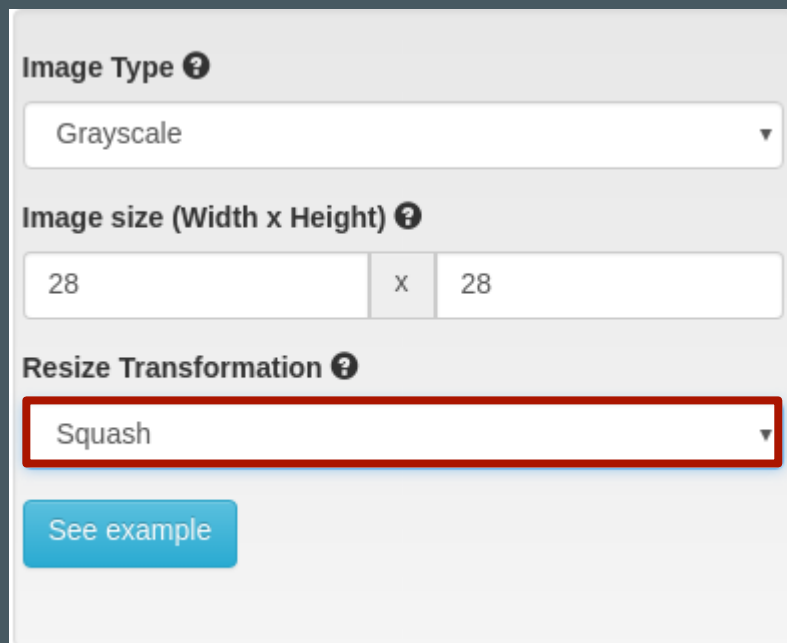
Squash ▼

See example

Image Size:

- Since it is MNIST dataset, we set the image size to 28x28

# Add a new dataset



The screenshot shows a configuration panel for adding a new dataset. It contains three main sections: 'Image Type' with a dropdown menu set to 'Grayscale'; 'Image size (Width x Height)' with two input fields both containing '28' and a separator 'x'; and 'Resize Transformation' with a dropdown menu set to 'Squash'. The 'Squash' option is highlighted with a red rectangular border. Below these sections is a blue button labeled 'See example'.

Image Type ?

Grayscale ▼

Image size (Width x Height) ?

28 x 28

Resize Transformation ?

Squash ▼

See example

## Resize Transformation

- Crop : traditional crop
- Squash : traditional squash
- Fill : keep the original ratio by filling the gap with a color
- Half crop, half fill : combination of crop and fill

# Add a new dataset

Use Image Folder

Use Text Files

Training Images ?

/home/<username>/mnist/train

Minimum samples per class ?

Maximum samples per class ?

% for validation ?

25

% for testing ?

0

☐ Separate validation images folder

☒ Separate test images folder

Test Images

/home/<username>/mnist/test

Minimum samples per class ?

Maximum samples per class ?

## Training Images:

- Train directory containing samples per class, "labels.txt" and "train.txt"

# Add a new dataset

Use Image Folder

Use Text Files

Training Images ?

/home/<username>/mnist/train

Minimum samples per class ?

Maximum samples per class ?

% for validation ?

25

% for testing ?

0

☐ Separate validation images folder

☒ Separate test images folder

Test Images

/home/<username>/mnist/test

Minimum samples per class ?

Maximum samples per class ?

## Minimum sample per class

- Ignore the class sample if the number of samples is **less** than the specified number
- Blank means ignore this feature



# Add a new dataset

Use Image Folder

Use Text Files

Training Images ?

/home/<username>/mnist/train

Minimum samples per class ?

Maximum samples per class ?

% for validation ?

25

% for testing ?

0

☐ Separate validation images folder

☒ Separate test images folder

Test Images

/home/<username>/mnist/test

Minimum samples per class ?

Maximum samples per class ?

## Maximum sample per class

- Ignore the class sample if the number of samples is **greater** than the specified number
- Blank means ignore this feature

# Add a new dataset

Use Image Folder

Use Text Files

Training Images ?

/home/<username>/mnist/train

Minimum samples per class ?

Maximum samples per class ?

% for validation ?

25

% for testing ?

0

☐ Separate validation images folder

☒ Separate test images folder

Test Images

/home/<username>/mnist/test

Minimum samples per class ?

Maximum samples per class ?

% for validation

- Proportion of the dataset for validation

# Add a new dataset

Use Image Folder   Use Text Files

**Training Images ?**

/home/<username>/mnist/train

Minimum samples per class ?   Maximum samples per class ?

% for validation ?   % for testing ?

25   0

☐ Separate validation images folder

☒ **Separate test images folder**

**Test Images**

/home/<username>/mnist/test

Minimum samples per class ?   Maximum samples per class ?

## Separate test images folder

- Since the downloaded MNIST dataset has the test images separated, this option has to be checked

# Add a new dataset

DB backend

LMDB

Image Encoding ?

PNG (lossless)

Group Name

Dataset Name

MNIST

Create

## DB backend

- LMDB : Great **performance**, but not good for large dataset
- HDF5 : Great for **large dataset**, but not good in term of performance

# Add a new dataset

DB backend

LMDB

Image Encoding ?

PNG (lossless)

Group Name

Dataset Name

MNIST

Create

## Image Encoding (Reduce image sizes)

- None : Raw image
- PNG : 100% quality with less file size
- JPEG : 90% quality with file size less than PNG

# Add a new dataset

**DB backend**  

LMDB

**Image Encoding ?**  

PNG (lossless)

**Group Name**

**Dataset Name**  

MNIST

Create

Group Name (optional)

- This is just a reference name to group a bunch of datasets

# Add a new dataset

DB backend

LMDB ▼

Image Encoding ?

PNG (lossless) ▼

Group Name

Dataset Name

MNIST

Create

## Dataset Name (required)


- This is just a reference name to be called in importing a dataset

# Add a new dataset

DIGITS

Image Classification Dataset

lukash (Logout) Info About

MNIST  
Owner: lukash

Clone Job Delete Job

Job Information

**Job Directory**  
/home/lukash/.digits\_jobs/20170702-193615-48b1

**Image Dimensions**  
28x28 (Width x Height)

**Image Type**  
Grayscale

**Resize Transformation**  
Squash

**DB Backend**  
Imdb

**Image Encoding**  
png

**DB Compression**  
none

**Dataset size**  
0 B

Job Status Done

- Initialized at 07:36:15 PM (1 second)
- Running at 07:36:16 PM (52 seconds)
- Done at 07:37:08 PM (Total: 52 seconds)

Parse Folder (train/val) Done

Parse Folder (test) Done

Create DB (train) Done

Create DB (val) Done

Create DB (test) Done

Notes



# Add a new model

## Home

2/2 GPUs available

No Jobs Running

[Datasets \(2\)](#)

[Models \(0\)](#)

[Pretrained Models \(1\)](#)

Group Jobs: ☒

New Model

[Images](#) ▾

Delete

Group



Filter



name

extension

framework

status

elapsed

submitted ^

No Models

# Add a new model

## Home

2/2 GPUs available

No Jobs Running

[Datasets \(2\)](#)[Models \(0\)](#)[Pretrained Models \(1\)](#)

Group Jobs: ☒

New Model

Images ▾

Classification

Object Detection

Other

Processing

Segmentation

DeleteGroup

nameextensionframeworkstat

No Models

# Add a new model

Select Dataset ?

MNIST

Cifar10

**MNIST**

Done 07:37:08 PM

**Image Size**  
28x28

**Image Type**  
GRAYSCALE

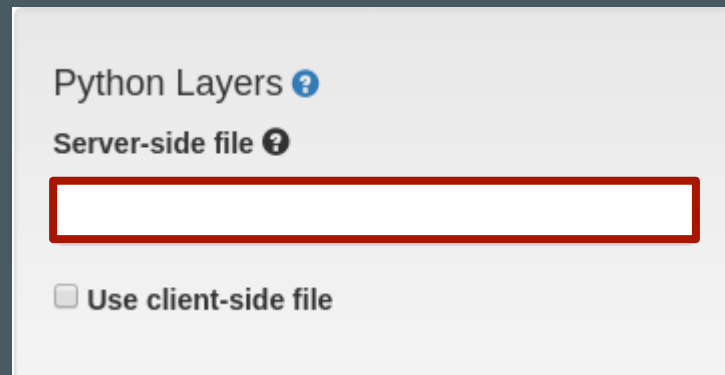
**DB backend**  
lmbd

**Create DB (train)**  
45002 images

**Create DB (val)**  
14998 images

**Create DB (test)**  
10000 images

# Add a new model



Python Layers ?

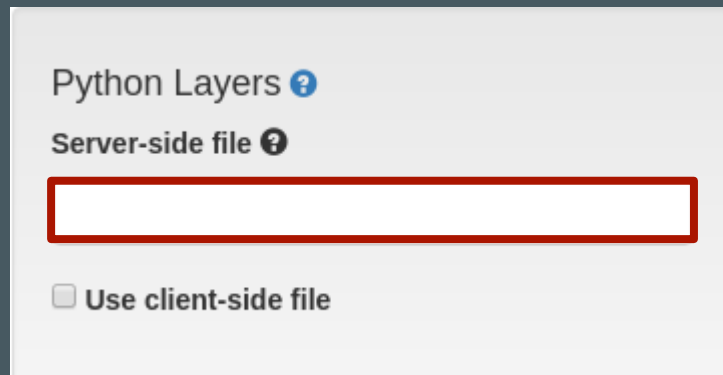
Server-side file ?

☐ Use client-side file

## Python Layers

- DIGITS support custom python layer using "Caffe"
- For now, leave it blank to keep things simple

# Add a new model



Python Layers ?

Server-side file ?

☐ Use client-side file

## Python Layers

- DIGITS support custom python layer using "Caffe"
- For now, leave it blank to keep things simple

# Add a new model

**Solver Options**

**Training epochs** ?

**Snapshot interval (in epochs)** ?

**Validation interval (in epochs)** ?

**Random seed** ?

## Training epochs

- How many passes through the training data

# Add a new model

**Solver Options**

**Training epochs** ?

  
**Snapshot interval (in epochs)** ?

1

**Validation interval (in epochs)** ?

1

**Random seed** ?

[none]

## Snapshot interval

- How frequent the model takes snapshots

# Add a new model

**Solver Options**

**Training epochs** ?

**Snapshot interval (in epochs)** ?

**Validation interval (in epochs)** ?

**Random seed** ?

## Validation interval

- How frequent the model calculates the accuracy



# Add a new model

**Solver Options**

**Training epochs** ?

  
**Snapshot interval (in epochs)** ?  
**Validation interval (in epochs)** ?  
**Random seed** ?

## Random Seed

- Determine the randomness of the dataset
- blank : the dataset will mostly different for the next training step
- any-number : the dataset will be

# Add a new model

**Batch size** ⓘ multiples allowed

[network defaults]

**Batch Accumulation** ⓘ

**Solver type** ⓘ

Stochastic gradient descent (SGD) ▼

**Base Learning Rate** ⓘ multiples allowed

0.01

☐ Show advanced learning rate options

## Batch size

- How many images to process at once
- blank : use the network configuration

# Add a new model

**Batch size** ⓘ multiples allowed

[network defaults]

**Batch Accumulation** ⓘ

**Solver type** ⓘ

Stochastic gradient descent (SGD) ▼

**Base Learning Rate** ⓘ multiples allowed

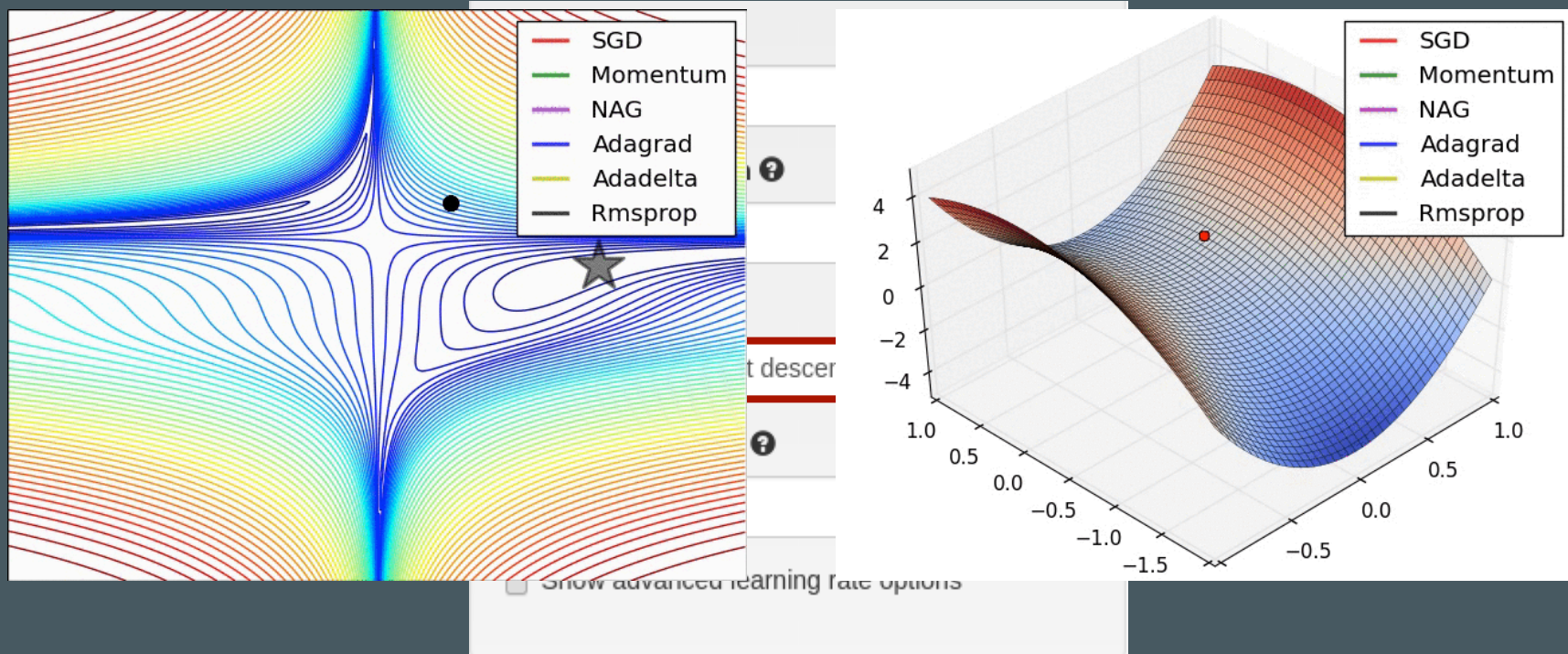
0.01

☐ Show advanced learning rate options

## Batch Accumulation

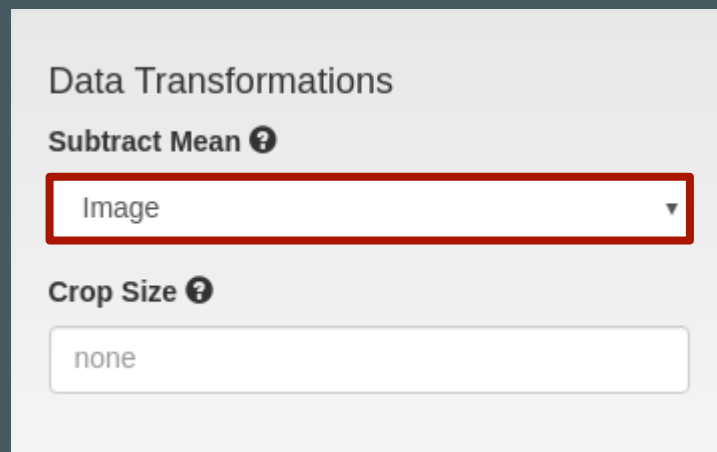
- How frequent the model calls the solver to readjust the variables

# Add a new model



Reference: [http://sebastianruder.com/content/images/2016/09/saddle\\_point\\_evaluation\\_optimizers.gif](http://sebastianruder.com/content/images/2016/09/saddle_point_evaluation_optimizers.gif)

# Add a new model



Data Transformations

**Subtract Mean ?**

Image ▼

**Crop Size ?**

none

## Subtract Mean

- **Image** : subtracted from any input image you feed to the neural network
- **Pixel** : subtract the "same" mean pixel value from all pixels of the input to the neural network

# Add a new model

Data Transformations

**Subtract Mean ?**

Image ▼

**Crop Size ?**

none

## Crop Size

- Remove some pixels from the images before they are given to the neural network
- Very good in autoencoder neural network

# Add a new model

Standard Networks			
Previous Networks			
Pretrained Networks			
Custom Network			
Caffe			
Network	Details	Intended image size	
<input checked="" type="radio"/> LeNet	<a href="#">Original paper</a> [1998]	28x28 (gray)	<a href="#">Customize</a>
<input type="radio"/> AlexNet	<a href="#">Original paper</a> [2012]	256x256	
<input type="radio"/> GoogLeNet	<a href="#">Original paper</a> [2014]	256x256	

# Add a new model

Use this many GPUs (next available)

1

or

Select which GPU[s] you would like to use ?

#0 - Tesla K20m (4.63 GB memory)  
#1 - Tesla K20m (4.63 GB memory)

Group Name ?

Model Name ?

Create

How many GPUs

- Since we only requested 1 GPU, we have to put 1 GPU



# Add a new model

Use this many GPUs (next available)

1

or

Select which GPU[s] you would like to use ?

#0 - Tesla K20m (4.63 GB memory)

#1 - Tesla K20m (4.63 GB memory)

Group Name ?

Model Name ?

Create

Select which GPU

- It's listed two GPUs because the two GPUs are in one node

# Add a new model

Use this many GPUs (next available)

1

or

Select which GPU[s] you would like to use ?

#0 - Tesla K20m (4.63 GB memory)  
#1 - Tesla K20m (4.63 GB memory)

Group Name ?

Model Name ?

Create

Group Name (optional)

- This is just a reference name to group a bunch of models

# Add a new model

Use this many GPUs (next available)

or

Select which GPU[s] you would like to use ?

#0 - Tesla K20m (4.63 GB memory)  
#1 - Tesla K20m (4.63 GB memory)

Group Name ?

Model Name ?

LeNet

Create

## Model Name (required)

- This is just a reference name to be called in importing a model

# Train

Use this many GPUs (next available)

1

or

Select which GPU[s] you would like to use ?

#0 - Tesla K20m (4.63 GB memory)  
#1 - Tesla K20m (4.63 GB memory)

Group Name ?

Model Name ?

LeNet

Create