The Niels Bohr Institute

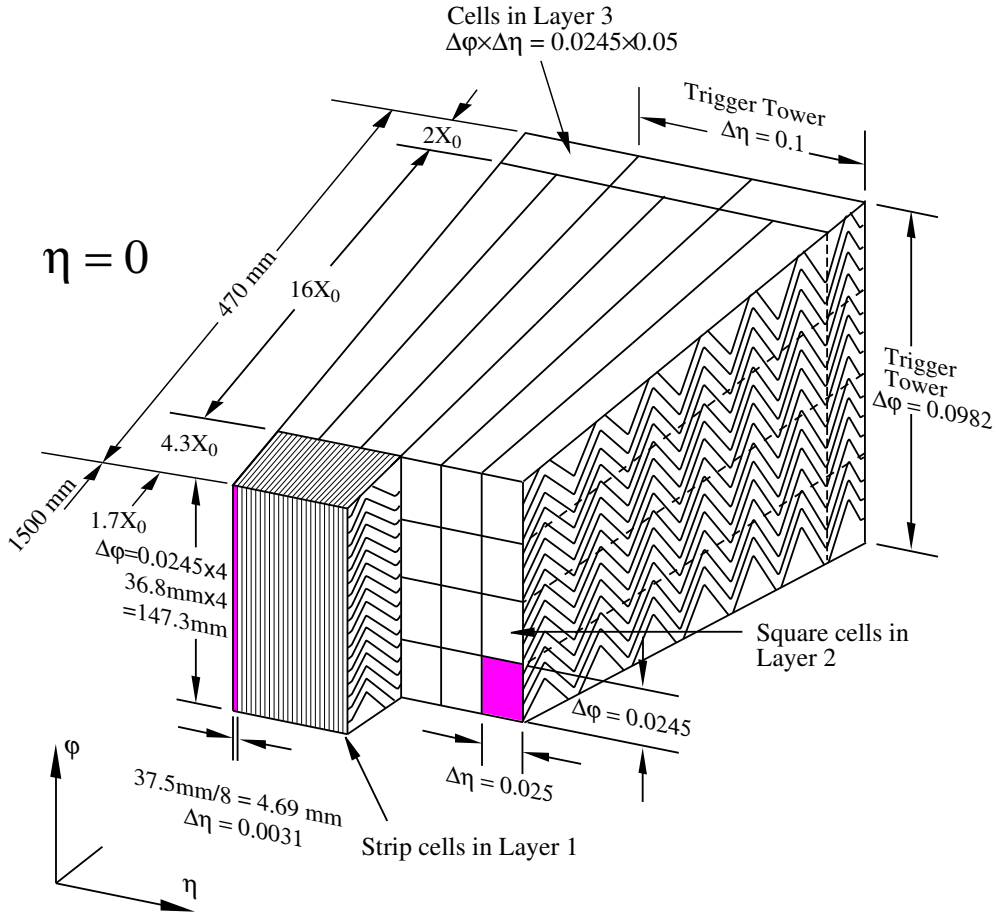
Energy reconstruction of electrons and photons using convolutional neural networks

Frederik G. Faye

SUPERVISION: Troels C. Petersen

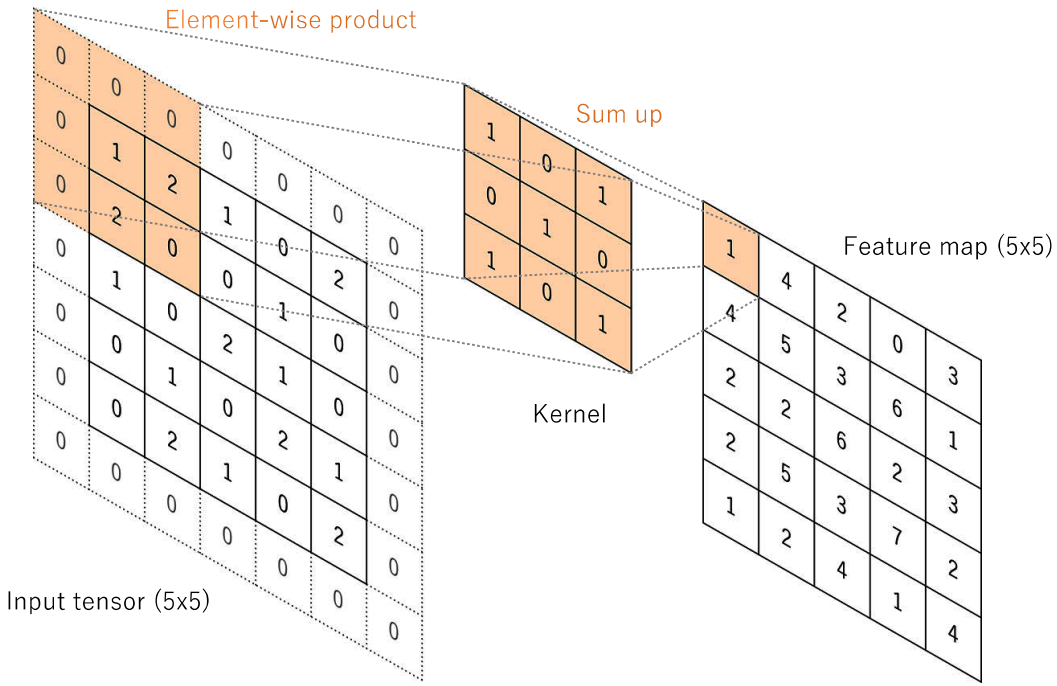
Spaatind, January 6, 2019

Basic idea

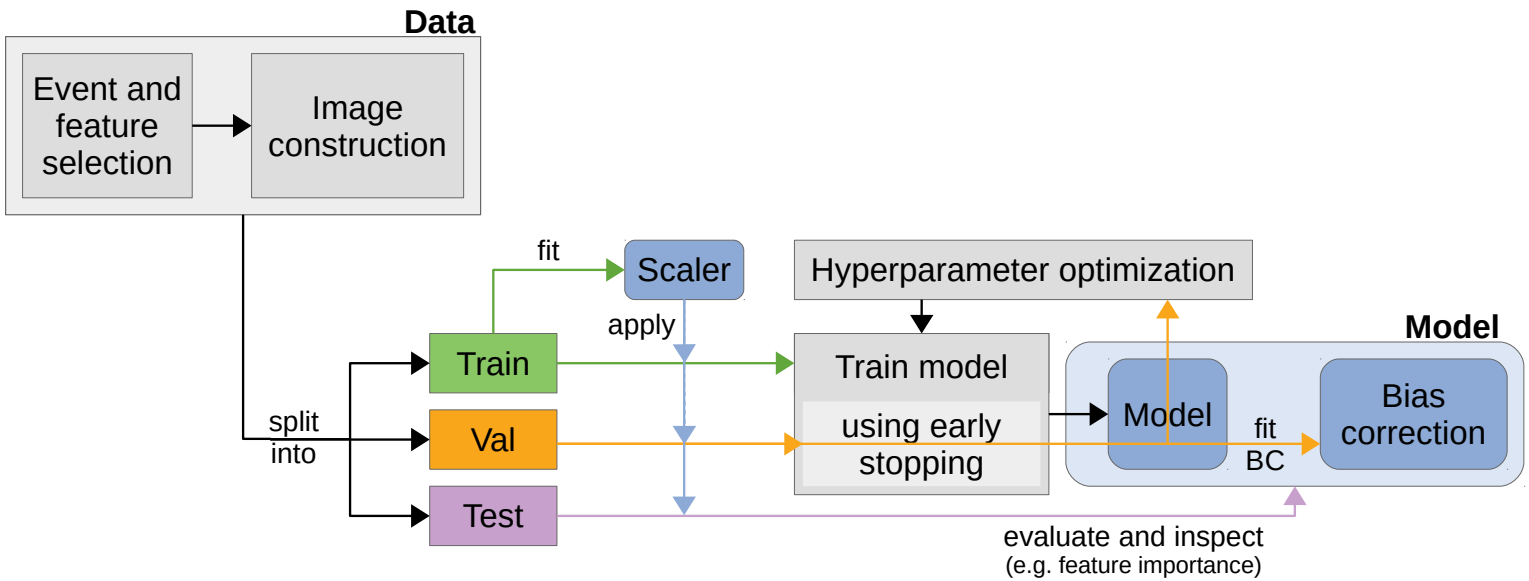


ECAL cells are like pixels in a multi-channel image,
so why not use CNNs?

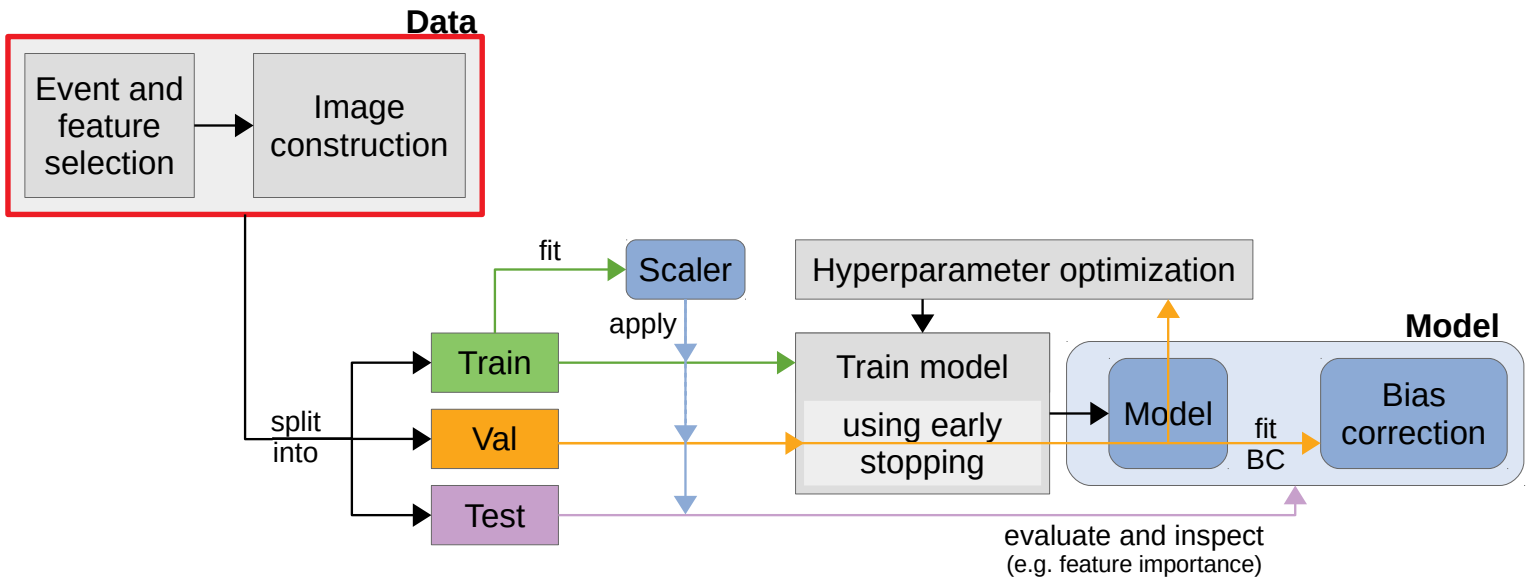
Convolutional neural networks



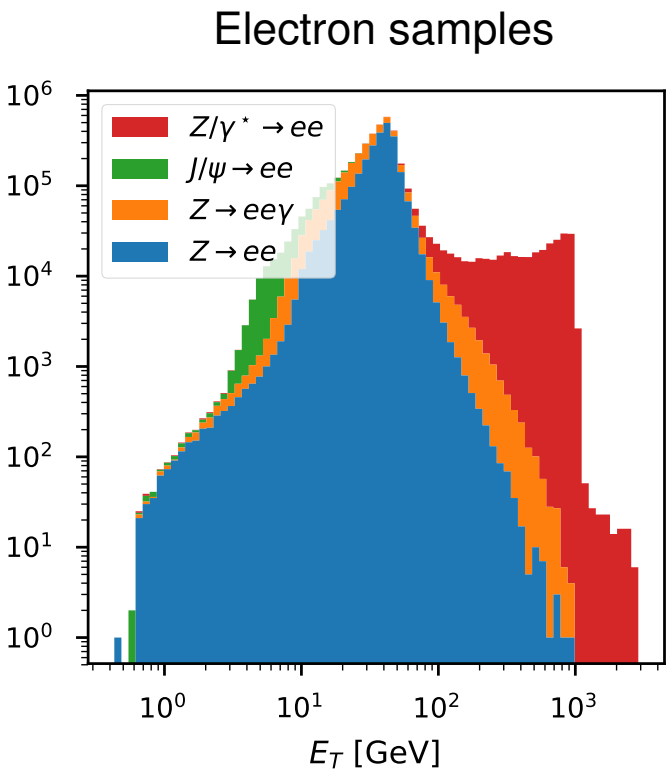
Pipeline

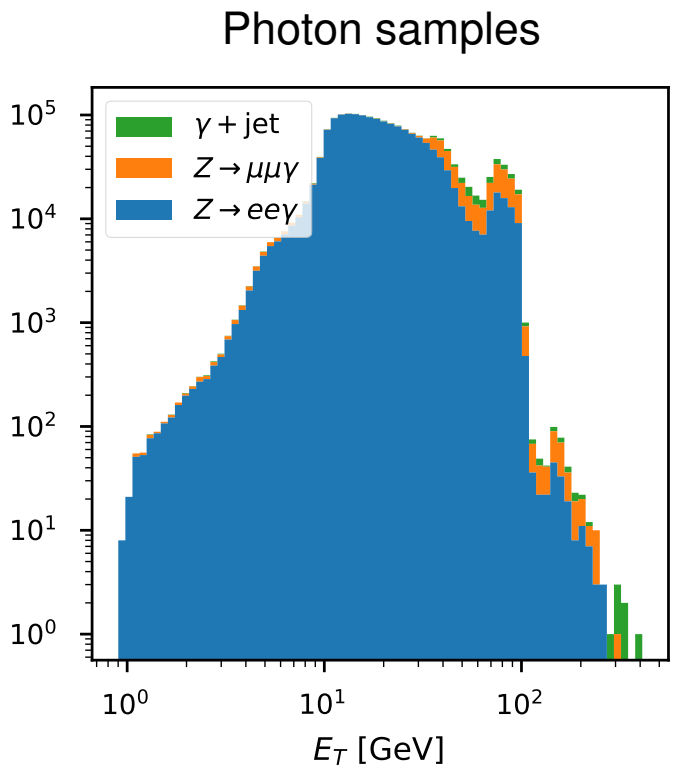

Data
Event and feature selection
Image construction
fit
Scaler
Hyperparameter optimization
apply
Model
Train
Train model
using early stopping
Model
Bias correction
fit BC
split into
Val
Test
evaluate and inspect
(e.g. feature importance)

Pipeline


Data
Event and feature selection
Image construction
fit
Scaler
Hyperparameter optimization
apply
Model
Train
Train model
using early stopping
Model
Bias correction
fit BC
split into
Val
Test
evaluate and inspect
(e.g. feature importance)

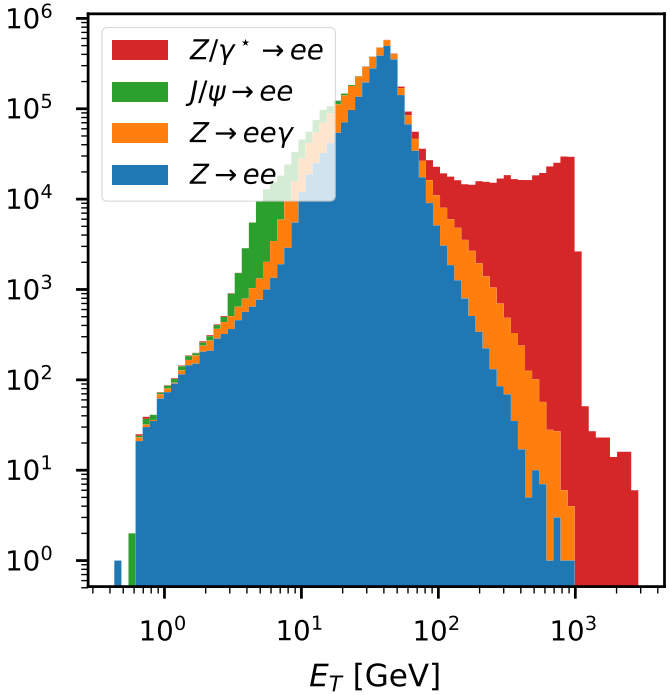
Simulation data



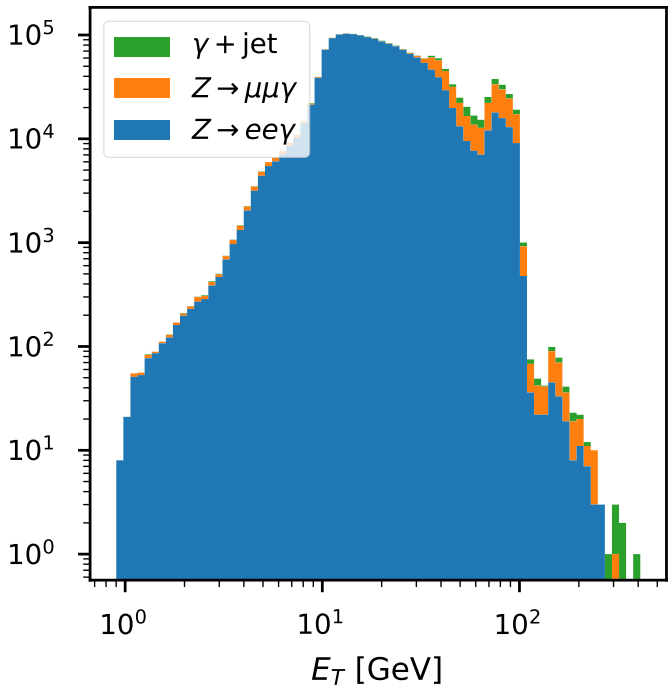


Simulation data

Electron samples



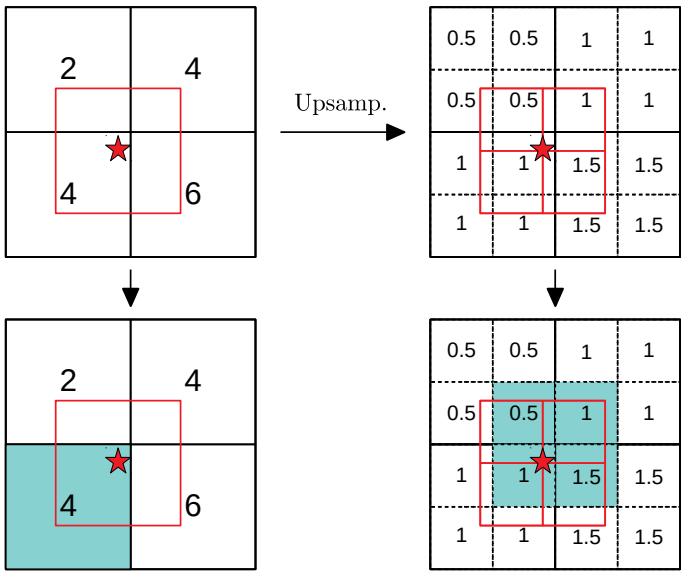
Photon samples



- Split into barrel, crack, endcap datasets
- $\sim 1 - 5$ M points in each dataset
- All truth-matched
- From 2016d with $\langle \mu \rangle \sim 37$

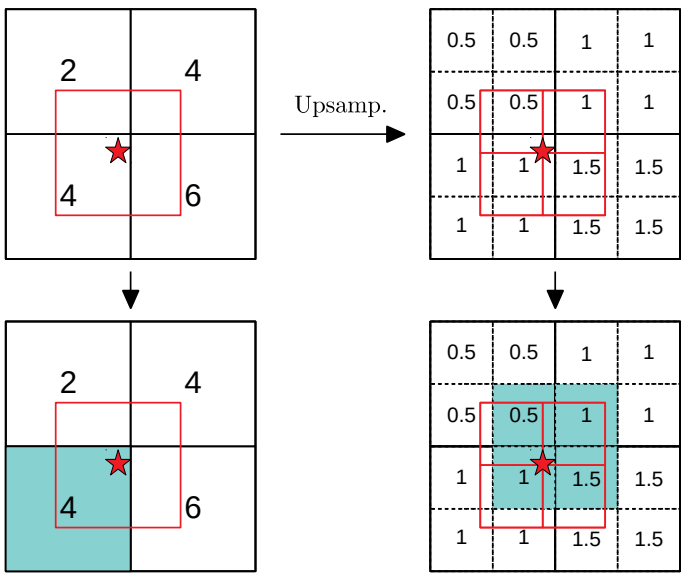
Simulation data

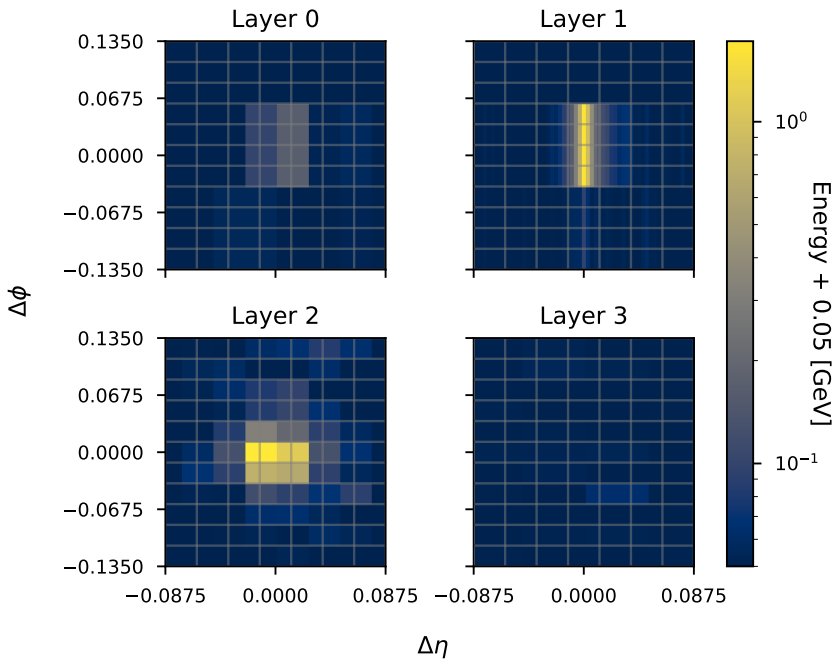
Images (code for producing here)



Simulation data

Images (code for producing here)

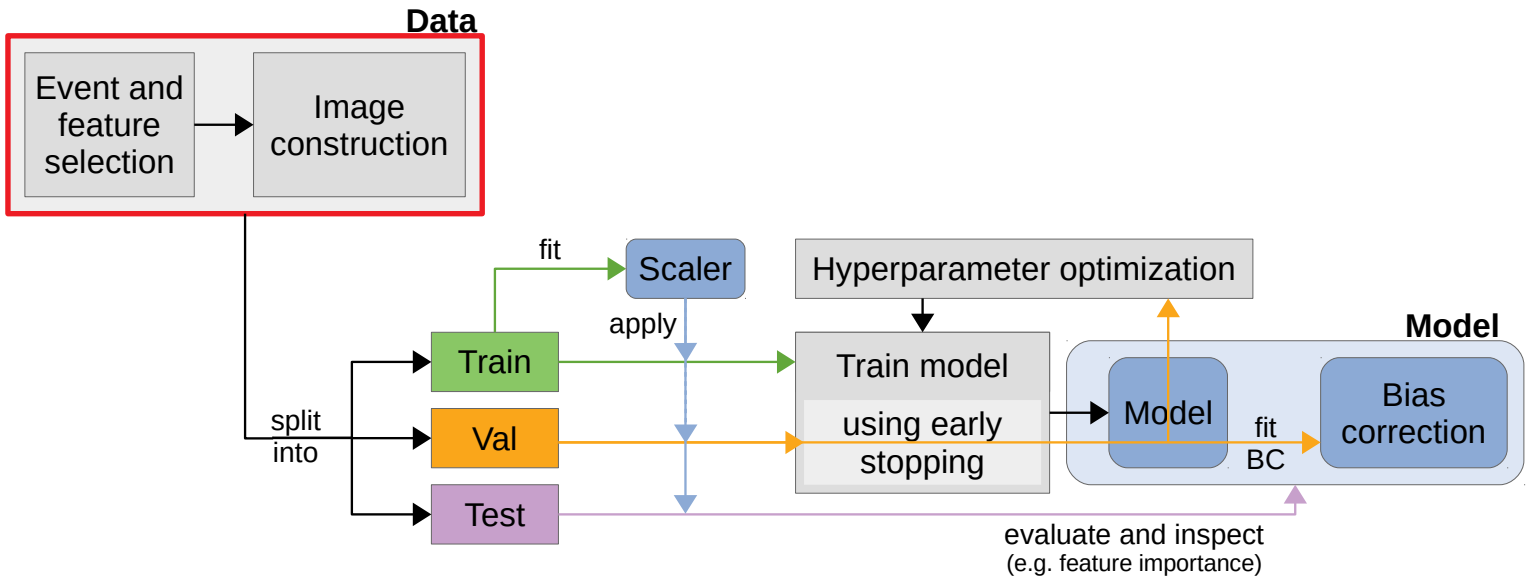


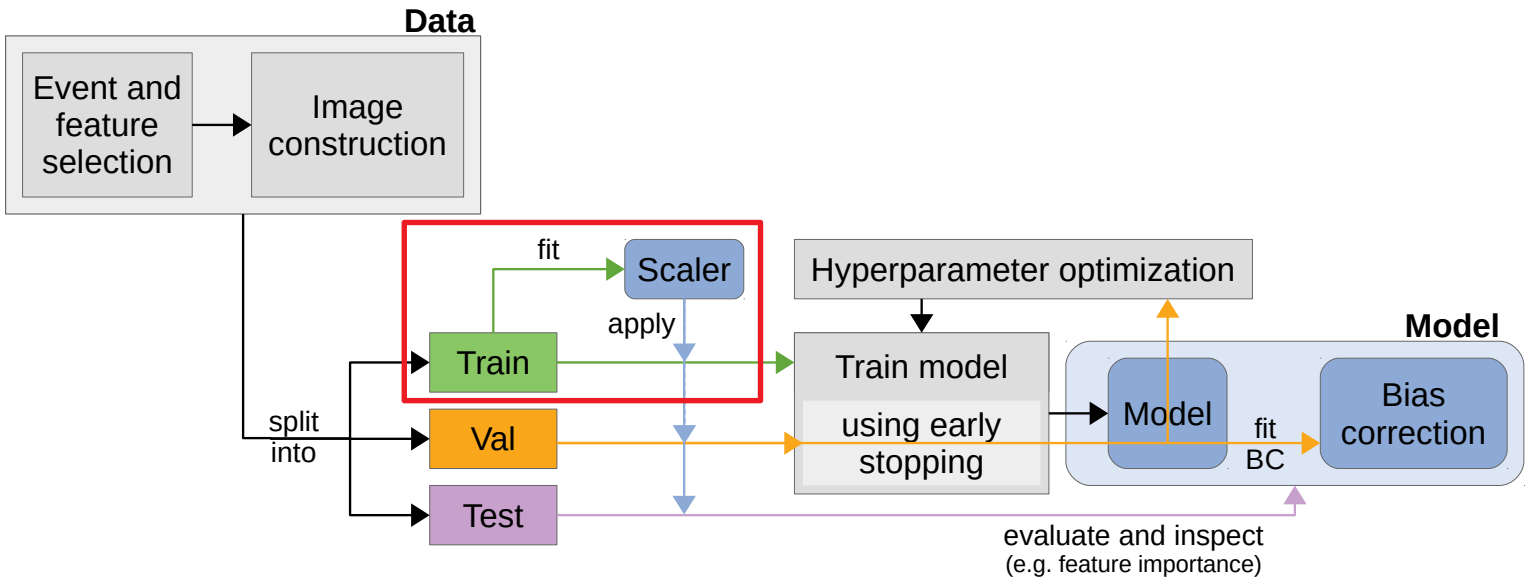


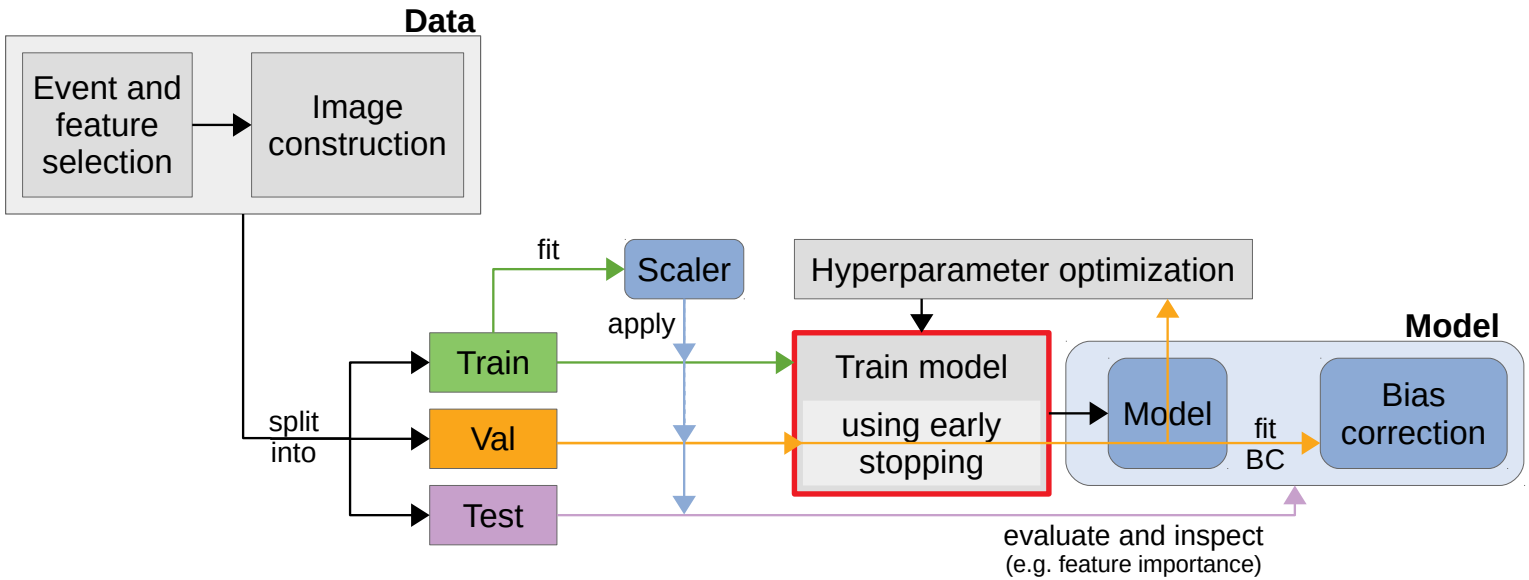
Simulation data

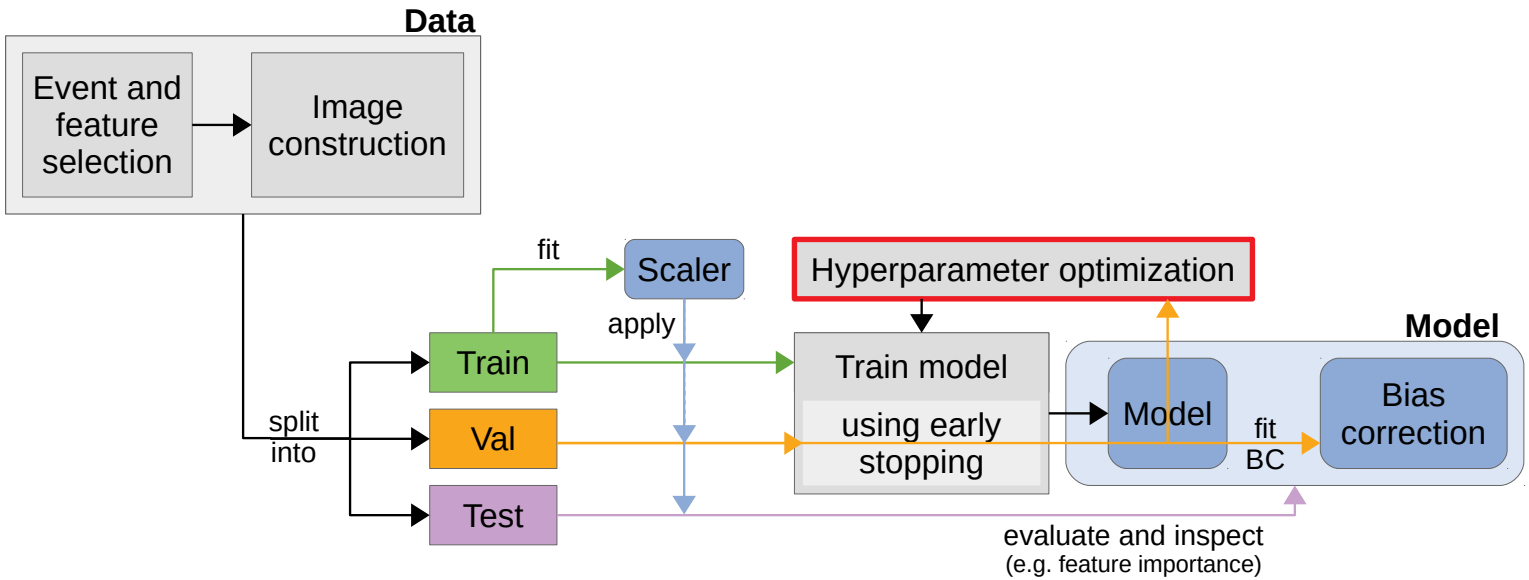
Scalars

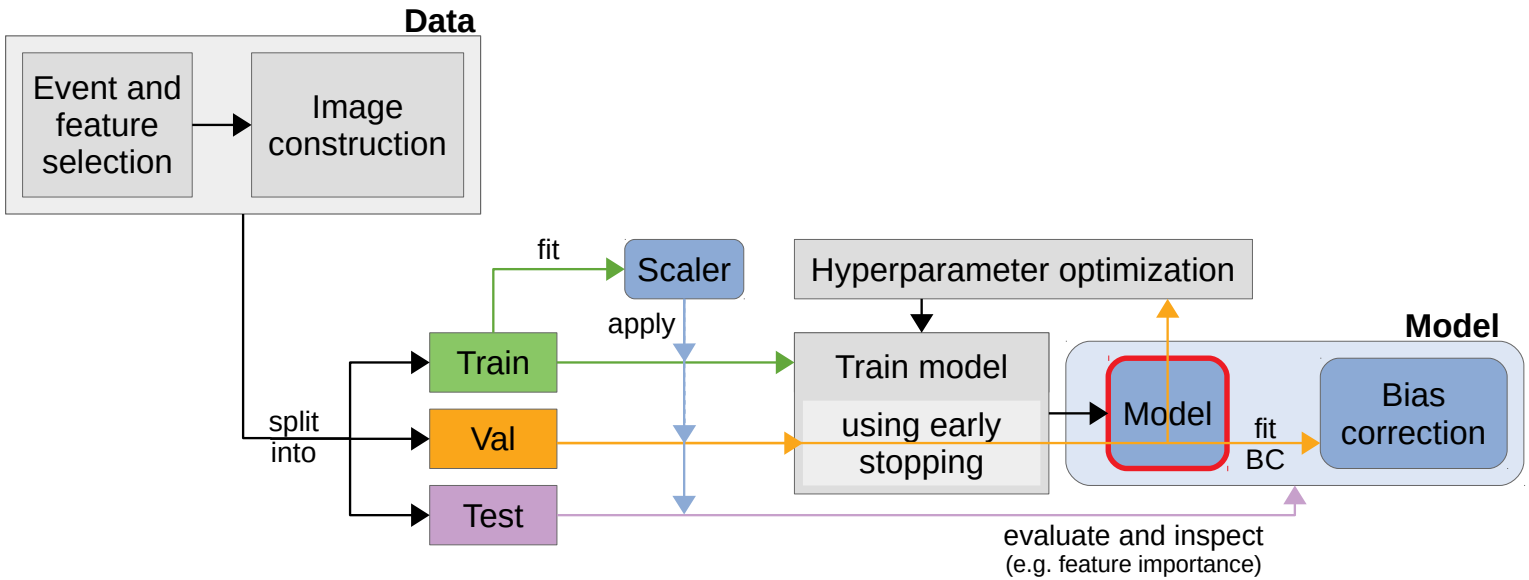
ECAL	E_{acc}	
	f_0	Fraction of energy in different ECAL layers
	R_{12}	
	E_{TG3}	
Context	η	
	η_{index}	
	$\langle\mu\rangle$	
	$n_{\text{vertexReco}}$	
	$\eta\text{modECAL}$	Relative position within cell
	poscs_2	
	$\Delta\phi_{\text{TH3}}$	
Tracking	$\Delta\phi_2^{\text{rescaled}}$	Difference in tracking-ECAL position estimate
	$\Delta\eta_2$	
	p_T^{track}	
	n_{tracks}	


Data
Event and feature selection
Image construction
fit
Scaler
Hyperparameter optimization
apply
Model
Train
Train model
using early stopping
Model
Bias correction
split into
Val
fit BC
Test
evaluate and inspect
(e.g. feature importance)

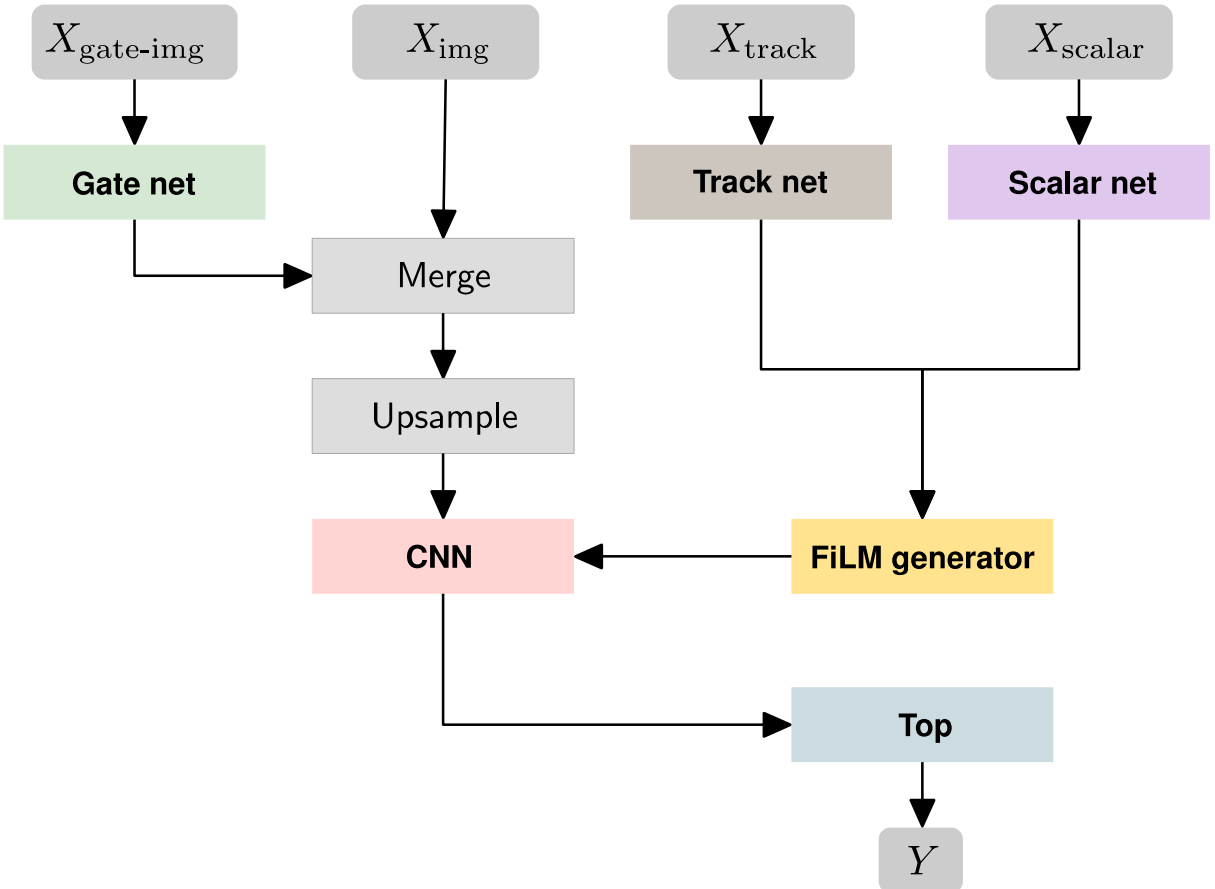

Data
Event and feature selection
Image construction
fit
Scaler
apply
Train
Val
Test
split into
Hyperparameter optimization
Train model using early stopping
Model
Model
Bias correction
fit BC
evaluate and inspect
(e.g. feature importance)


Data
Event and feature selection
Image construction
fit
Scaler
Hyperparameter optimization
apply
Model
Train
Train model
using early stopping
Model
Bias correction
split into
Val
fit BC
Test
evaluate and inspect
(e.g. feature importance)

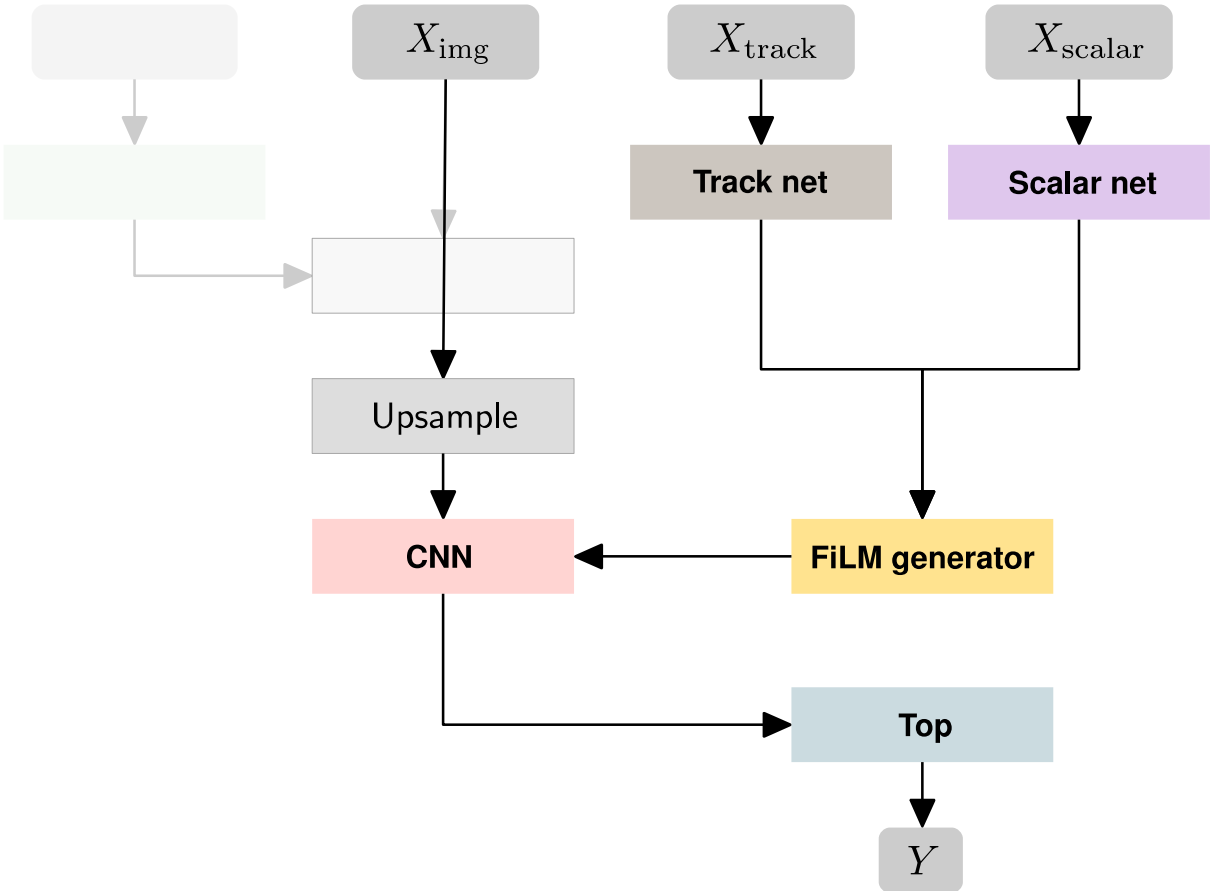

Data
Event and feature selection
Image construction
fit
Scaler
Hyperparameter optimization
apply
Model
Train
Train model
using early stopping
Model
Bias correction
split into
Val
fit BC
Test
evaluate and inspect
(e.g. feature importance)


Data
Event and feature selection
Image construction
fit
Scaler
Hyperparameter optimization
apply
Model
Train
Train model
using early stopping
Model
Bias correction
split into
Val
fit BC
Test
evaluate and inspect
(e.g. feature importance)

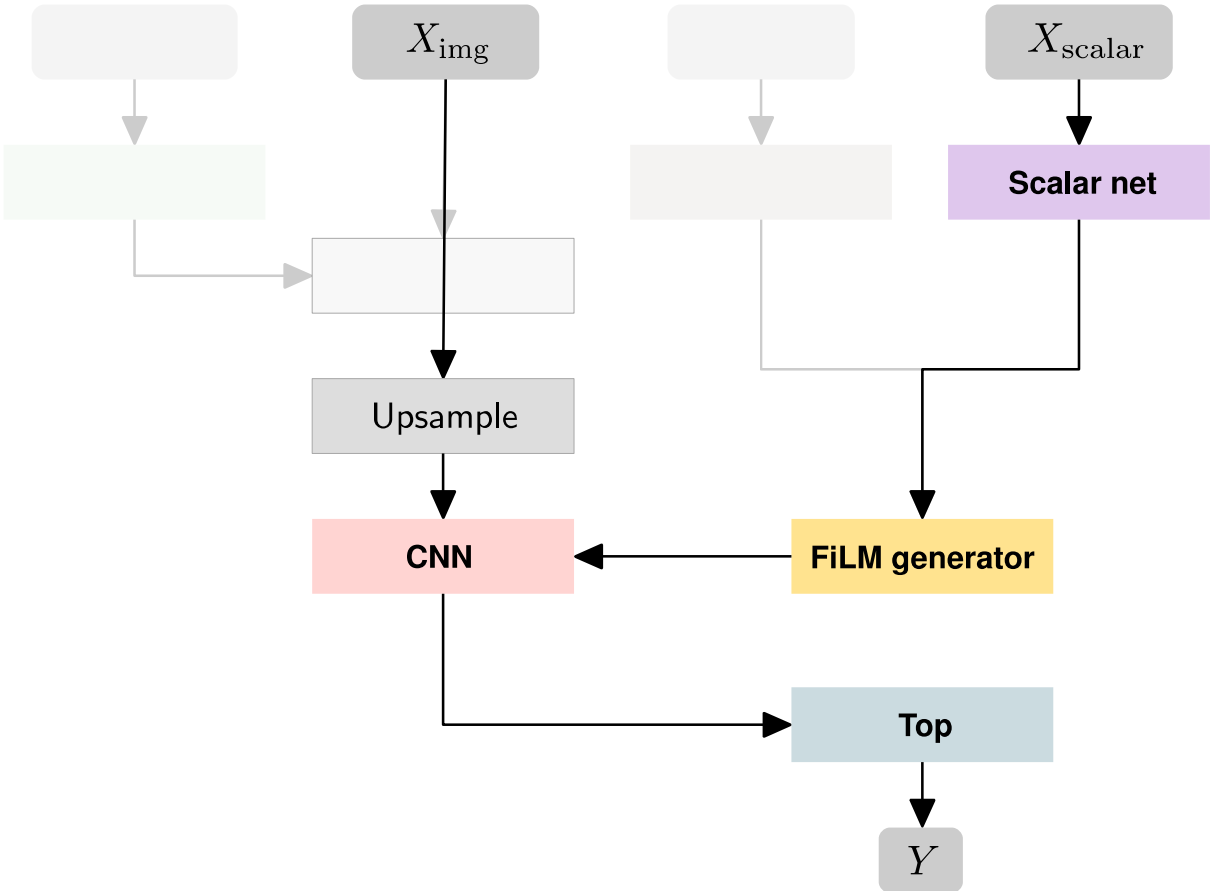
Model architecture



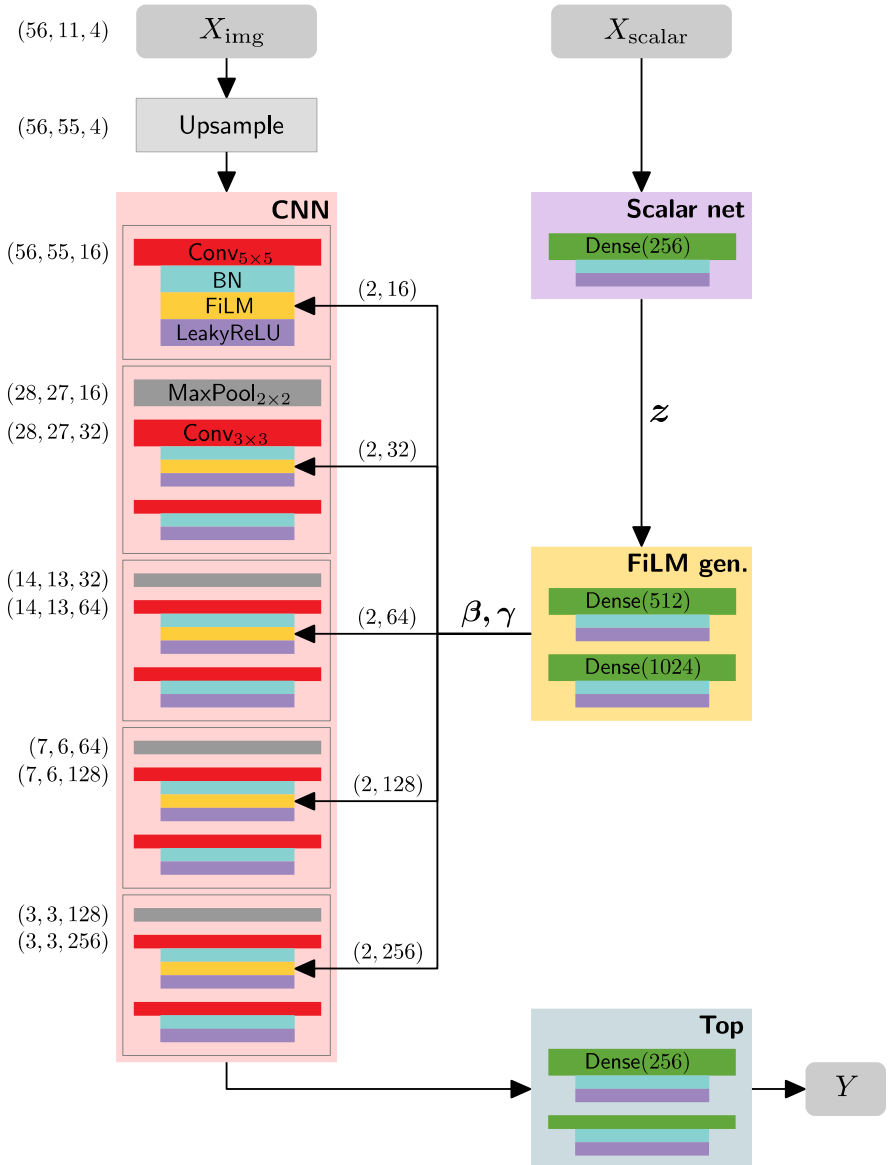
Model architecture



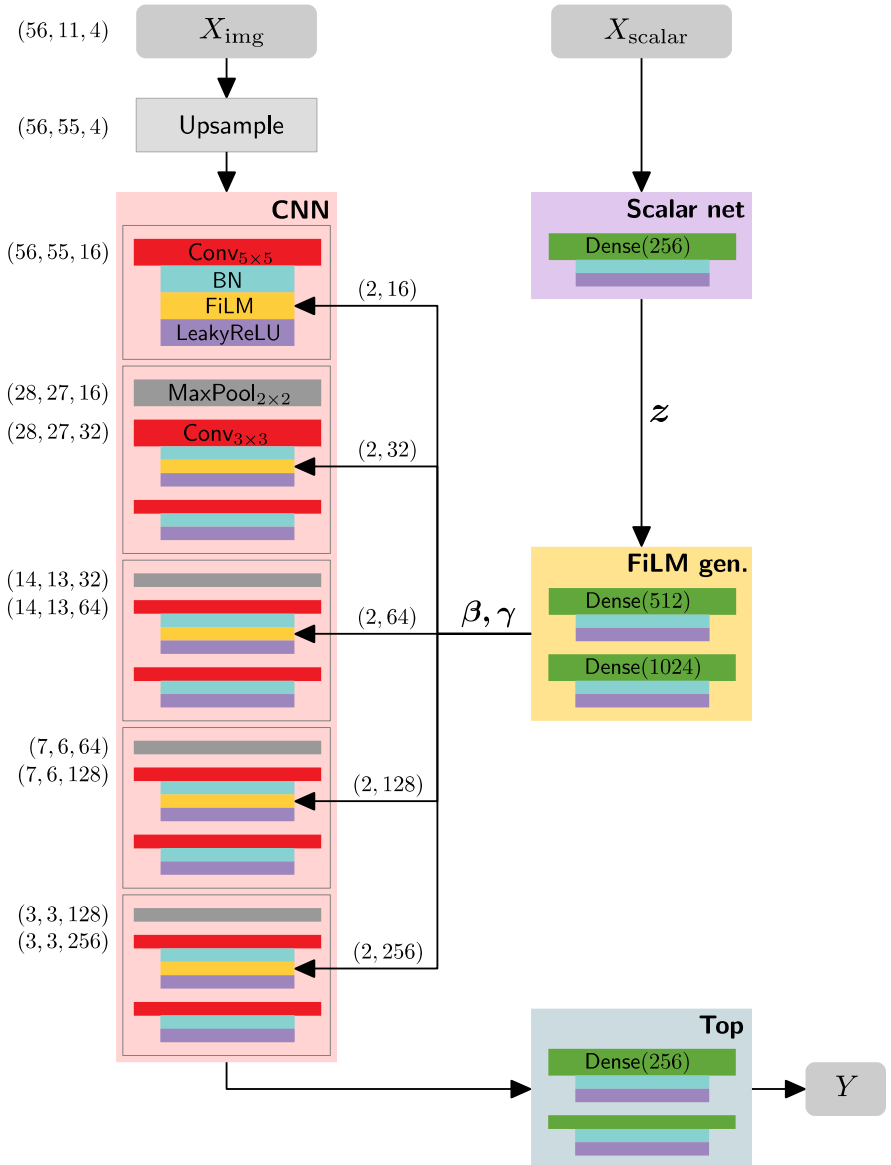
Model architecture



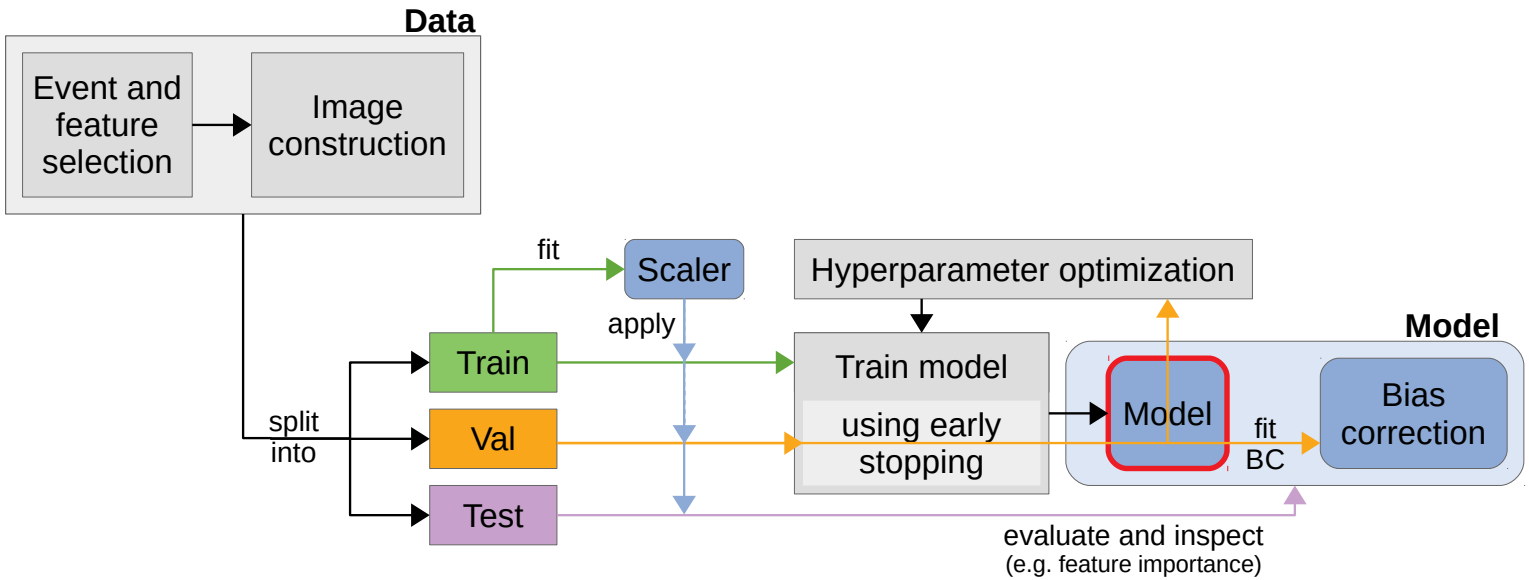
Model architecture

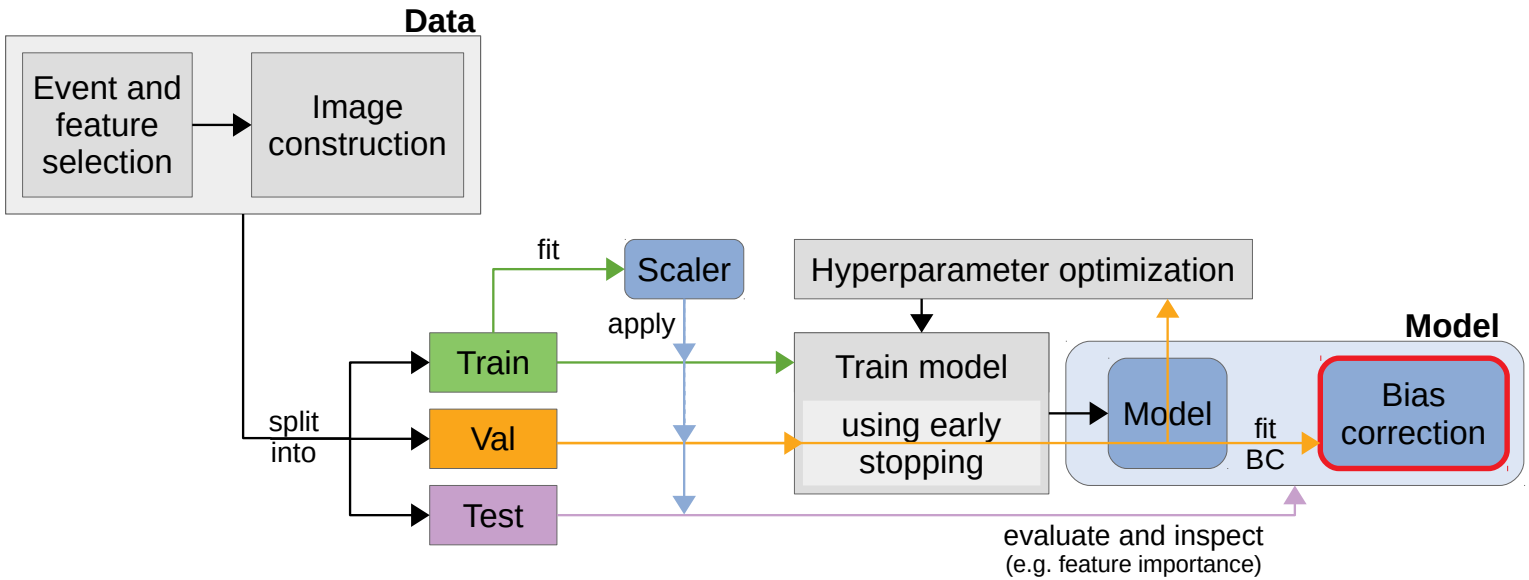


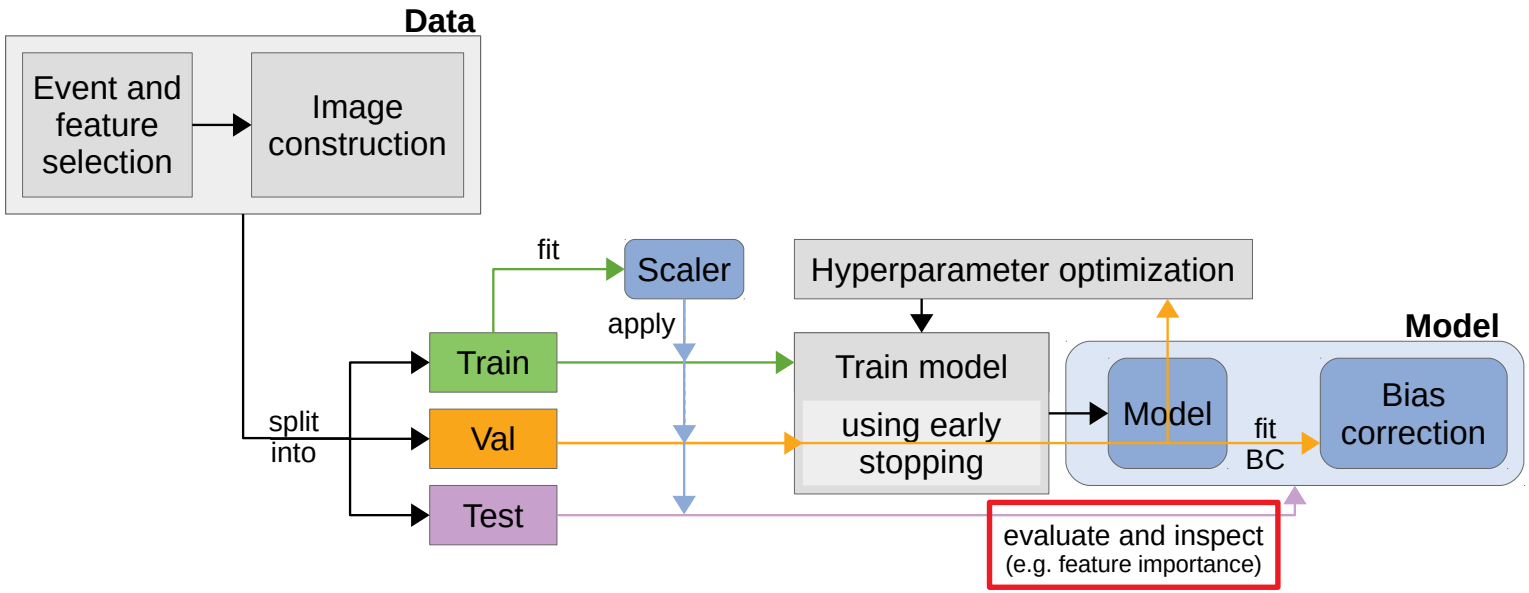
Model architecture



$$\text{FiLM}(x_i) = \gamma_i(\boldsymbol{z})x_i + \beta_i(\boldsymbol{z})$$


Data
Event and feature selection
Image construction
fit
Scaler
Hyperparameter optimization
apply
Model
Train
Train model
using early stopping
Model
Bias correction
split into
Val
fit BC
Test
evaluate and inspect
(e.g. feature importance)


Data
Event and feature selection
Image construction
fit
Scaler
Hyperparameter optimization
apply
Model
Train
Train model
using early stopping
Model
Bias correction
split into
Val
fit BC
Test
evaluate and inspect
(e.g. feature importance)


Data
Event and feature selection
Image construction
fit
Scaler
Hyperparameter optimization
apply
Model
Train
Train model
Model
Bias correction
split into
Val
using early stopping
fit BC
Test
evaluate and inspect
(e.g. feature importance)

Results

Metrics

Relative error:

$$\mathrm{RE} = \frac{E_{\mathrm{pred}} - E_{\mathrm{true}}}{E_{\mathrm{true}}}$$

Interquartile range (over distribution of RE):

$$\mathrm{IQR}(\mathrm{RE}) = Q_3(\mathrm{RE}) - Q_1(\mathrm{RE})$$

Relative improvement:

$$r\mathrm{IQR} = 1 - \mathrm{IQR}^{\mathrm{model}} / \mathrm{IQR}^{E\mathrm{calib}^{(\mathrm{BDT})}}$$

Results

Two experiments

	No E_T bins, barrel e	No η bins, low energy e/γ
E_T	$5 - 1000\ \mathrm{GeV}$	$5 - 100\ \mathrm{GeV}$
$\lvert\eta\rvert$	$0 - 1.3$	$0 - 2.5$ (2.4 for γ)
Trick	Predict $E_{\mathrm{true}}/E_{\mathrm{acc}}$	Add barrel and endcap images

Results

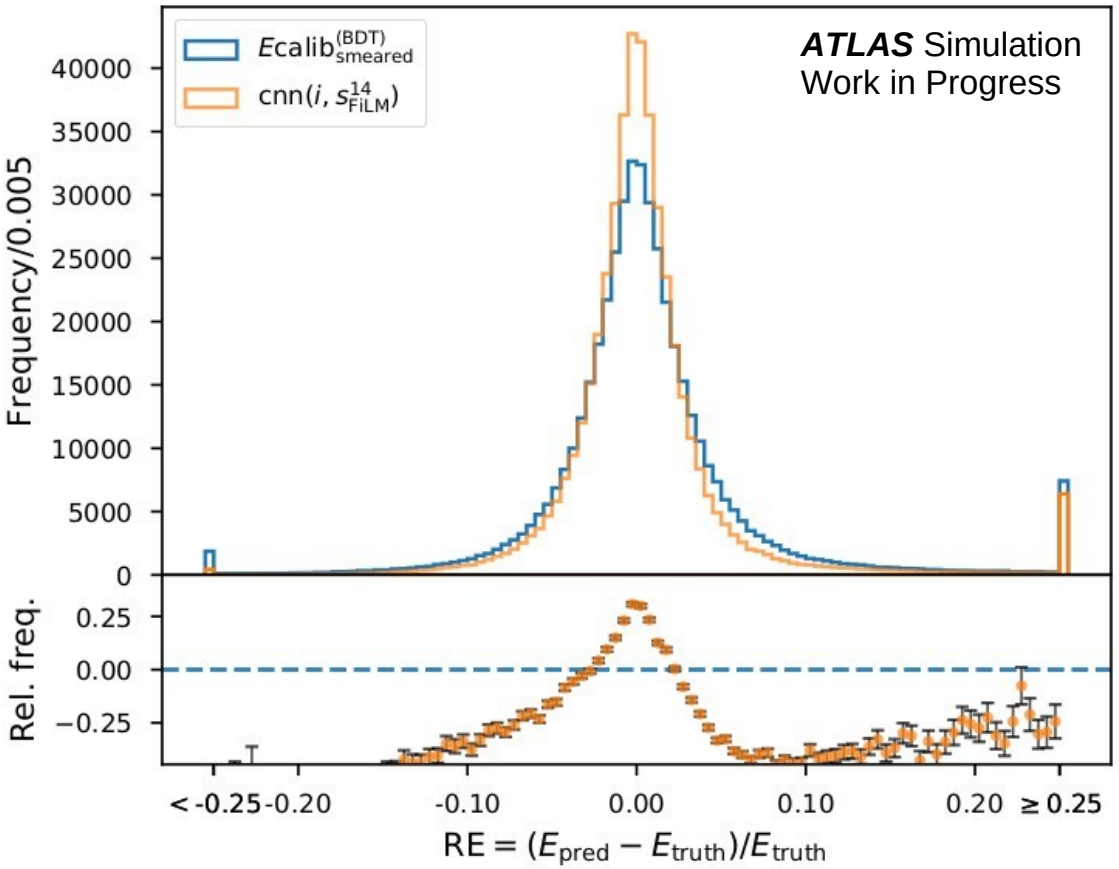
Two experiments

	No E_T bins, barrel e	No η bins, low energy e/γ
E_T	$5 - 1000\ \mathrm{GeV}$	$5 - 100\ \mathrm{GeV}$
$\lvert\eta\rvert$	$0 - 1.3$	$0 - 2.5$ (2.4 for γ)
Trick	Predict $E_{\mathrm{true}}/E_{\mathrm{acc}}$	Add barrel and endcap images

Both work, so plan is to merge them

Results

No E_T bins, barrel e

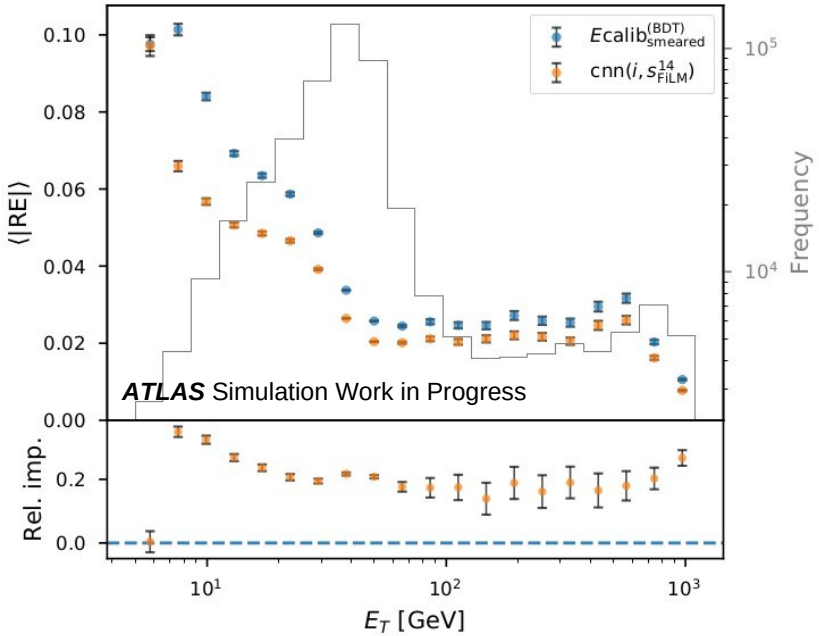


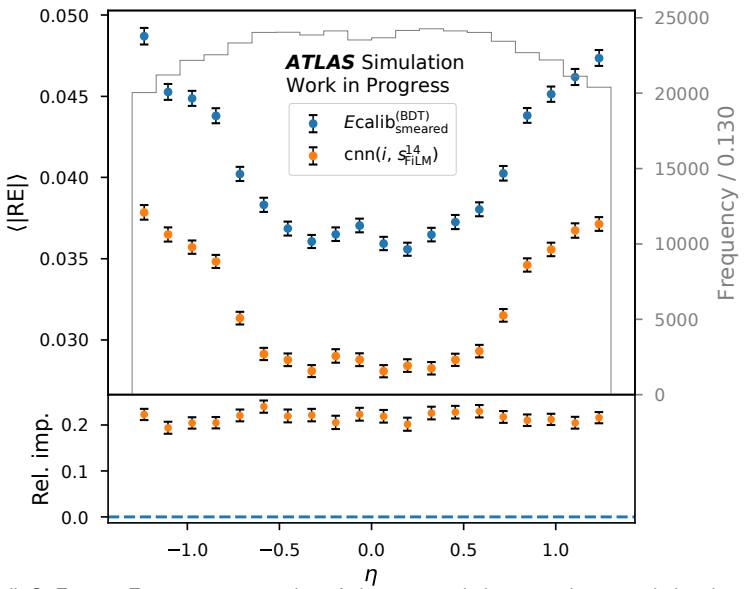
$$r\mathrm{IQR} = 24.0\%$$

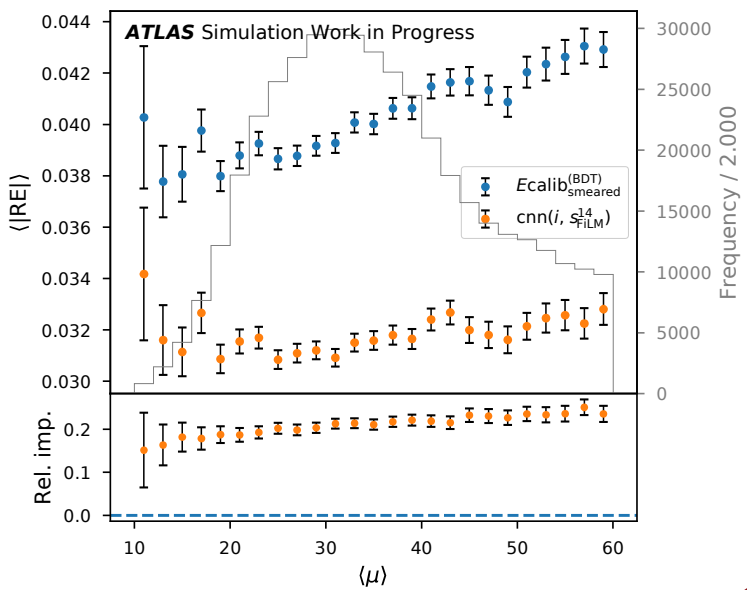
- This model predicts $E_{\mathrm{true}} / E_{\mathrm{acc}}$ (predicting E_{true} did not work here)
- $E\mathrm{calib}^{(\mathrm{BDT})}_{\mathrm{smeared}}$ is smeared $\rightarrow$ true $r\mathrm{IQR}$ is $1 - 2\%$ lower

Results

No E_T bins, barrel e





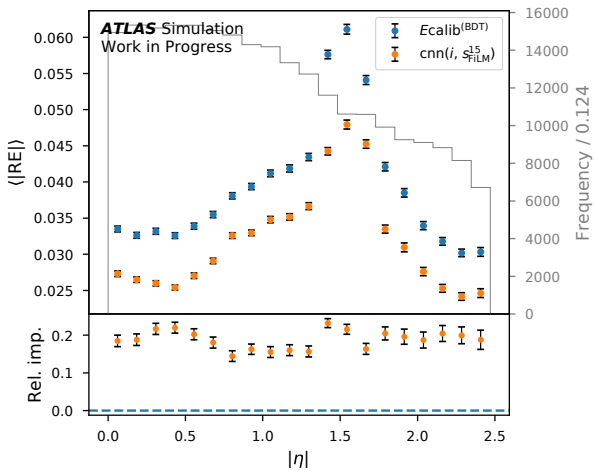


Results

No η bins, low energy e/γ

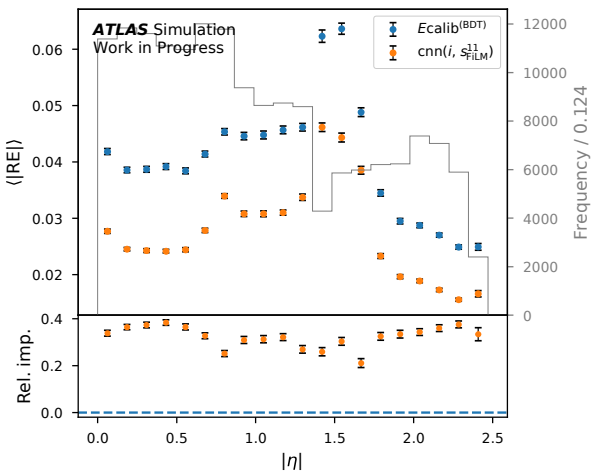
Single model trained on photons $< 100\ \mathrm{GeV}$

Electrons ($Z \to ee$)



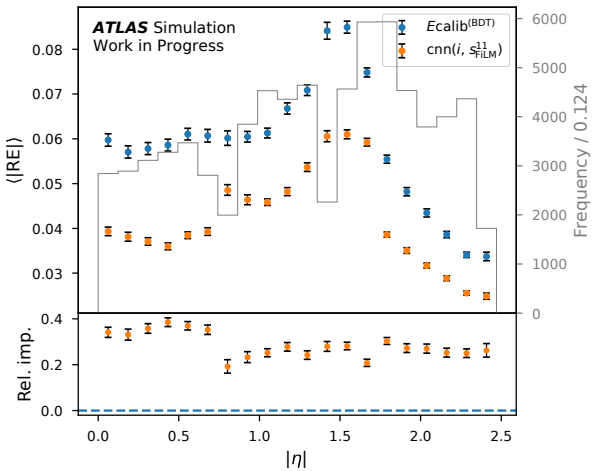
$rIQR = 20.8\%$

Unconverted



$rIQR = 29.4\%$

Converted

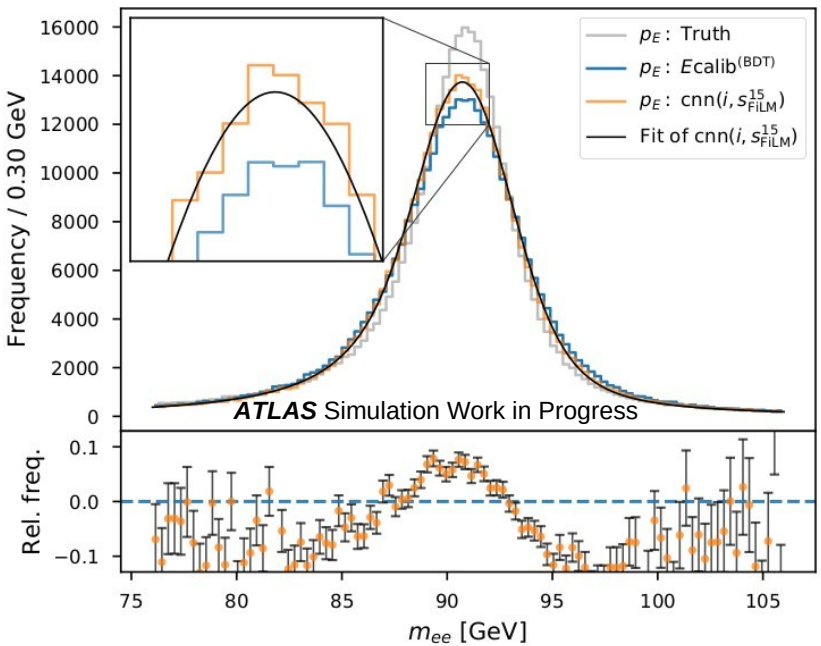


$rIQR = 19.9\%$

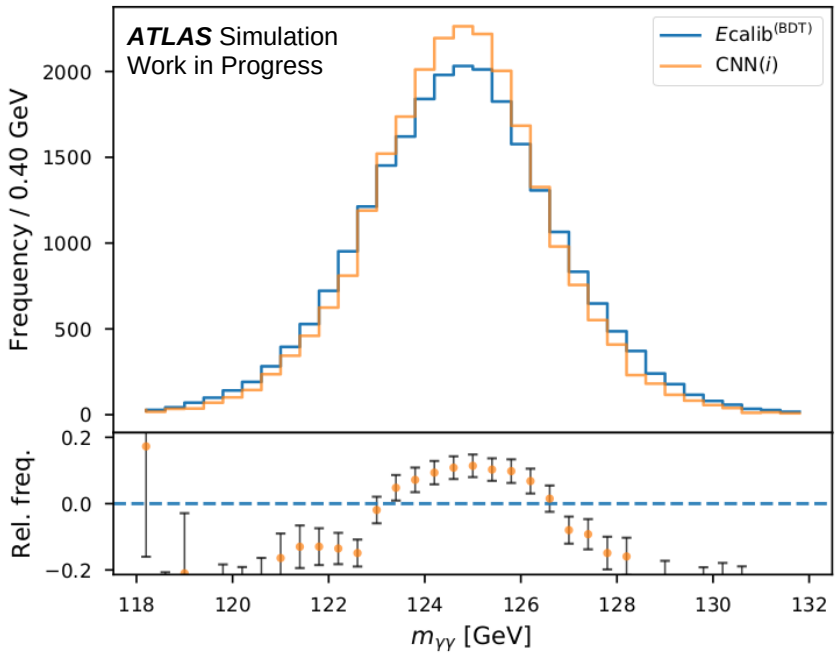
Results

No η bins, low energy e/γ

$Z \to ee$ (probe only)



$H \to \gamma\gamma$



$$\sigma_{\text{CB}}^{\text{Ecalib}} = 2.09(1)$$

$$\sigma_{\text{CB}}^{\text{cnn}} = 1.87(1)$$

$$1 - \sigma_{\text{CB}}^{\text{cnn}} / \sigma_{\text{CB}}^{\text{Ecalib}} = 10.5(7)\%$$

Very preliminary!

What's next?

- **Produce data** covering central region and whole E_T range
- **Train** three models (e, conv. and unconv. γ) each without E_T, η bins
- **Implement** in ATHENA (worked on by people in ML Forum)
- Apply currently used **MC** $\rightarrow$ **Data corrections** and test on Data
- Test if **MC** $\rightarrow$ **Data cell reweighting** improves performance in Data
- Find **minimum set of scalars** (guided by permutation imp. or SHAP)

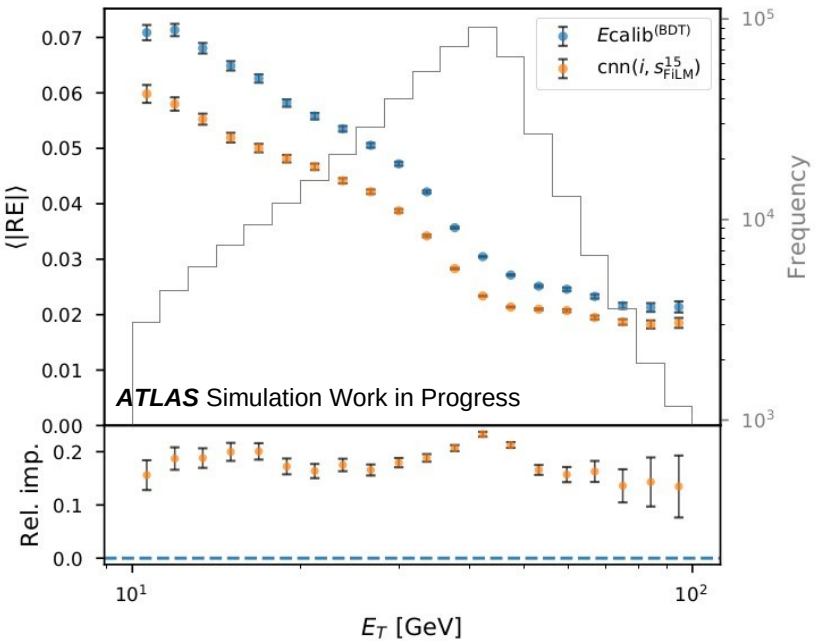
What's next?

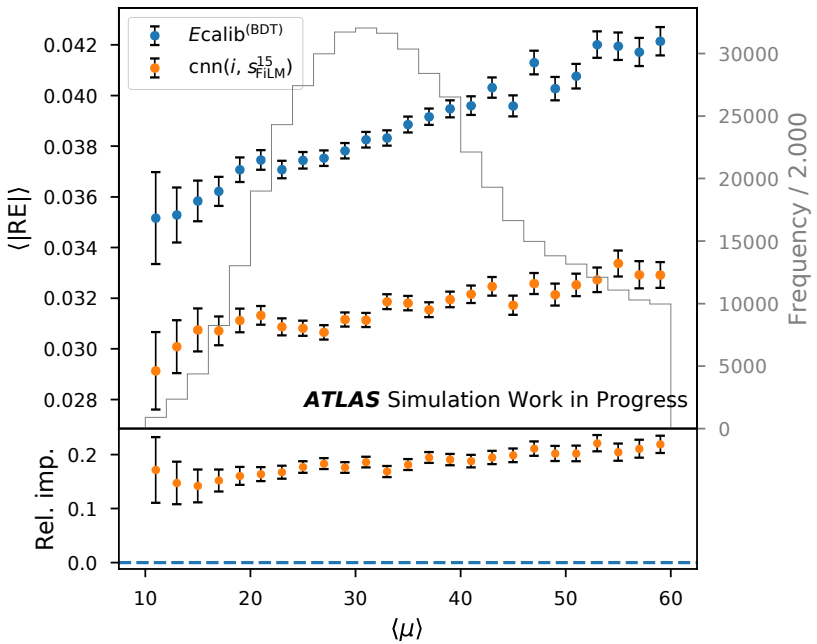
- **Produce data** covering central region and whole E_T range
- **Train** three models (e, conv. and unconv. γ) each without E_T, η bins
- **Implement** in ATHENA (worked on by people in ML Forum)
- Apply currently used **MC** $\rightarrow$ **Data corrections** and test on Data
- Test if **MC** $\rightarrow$ **Data cell reweighting** improves performance in Data
- Find **minimum set of scalars** (guided by permutation imp. or SHAP)

- Well-documented code is available at gitlab.com/ffaye/deepcalo (`pip install deepcalo`)
- Short **summary of recommendations** based on my thesis available here
- Internal note is on its way

Backup

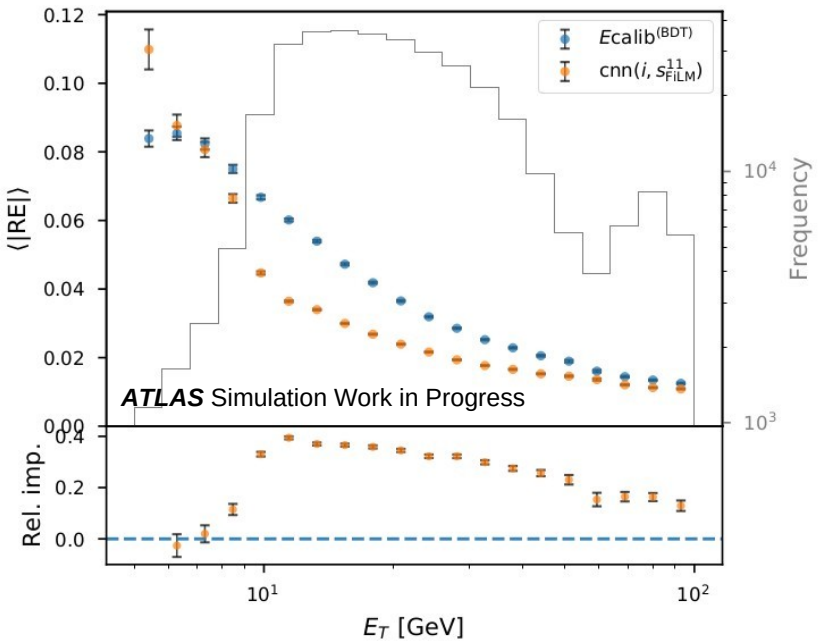
No η bins, $Z \rightarrow ee$ electrons

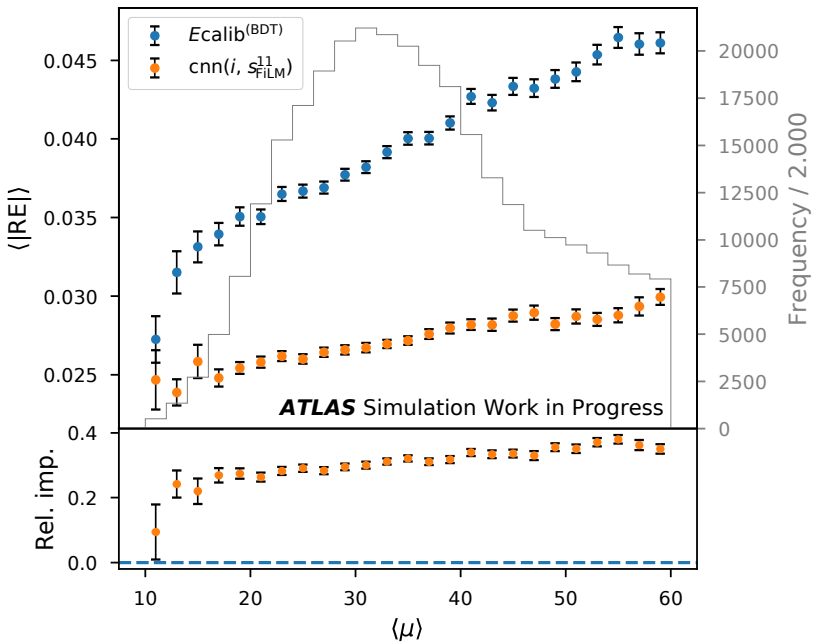




Backup

No η bins, unconverted electrons < 100 GeV





Backup

No η bins, converted electrons < 100 GeV