Super-resolution of photon calorimeter images using generative adversarial networks

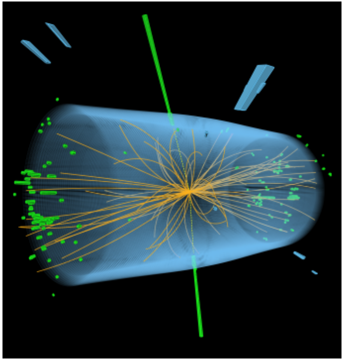
Florian Mausolf

Work in progress by J. Erdmann[a], A. van der Graaf[b], F. Mausolf[a] and O. Nackenhorst[b]

14th September 2022

[a]RWTH Aachen University,
[b]TU Dortmund University

- Photons are important physics objects at collider experiments
  - $H \to \gamma\gamma$: clean channel to study the Higgs boson
  - Investigation of electroweak interactions
  - Searches for new physics with photons, ...
- Signature: cluster of energy deposition in the EM calorimeter

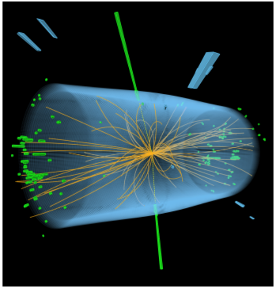


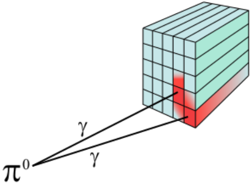
$H \to \gamma\gamma$ candicate event at CMS

cds.cern.ch/record/2736135/

Introduction

- Photons are important physics objects at collider experiments
  - $H \to \gamma\gamma$: clean channel to study the Higgs boson
  - Investigation of electroweak interactions
  - Searches for new physics with photons, ...
- Signature: cluster of energy deposition in the EM calorimeter
- Rejection of fake photons is a crucial and challenging task
  - Main source: highly collimated photons from high-energy $\pi^0 \to \gamma\gamma$ decays
  - High granularity calorimeter needed to resolve collimated photons

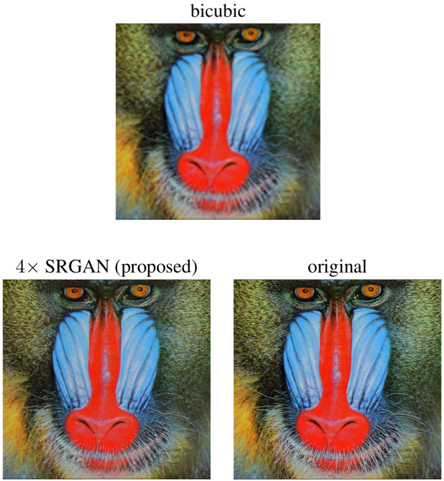




Boosted $\pi^0 \to \gamma\gamma$ decay

wikimedia.org/wiki/File:
Pion_Decay_Two_Photons_CMS/

- Single-image super-resolution (SR): estimation of a high resolution (HR) image from a single low resolution (LR) image
- Intensively studied in the field of image processing
- Has been applied to jet physics (arXiv: 2012.11944)
- State-of-the-art models based on deep CNNs
  - Trained on LR-HR image pairs
- GAN-based training setups particularly successfull in producing realistic SR images

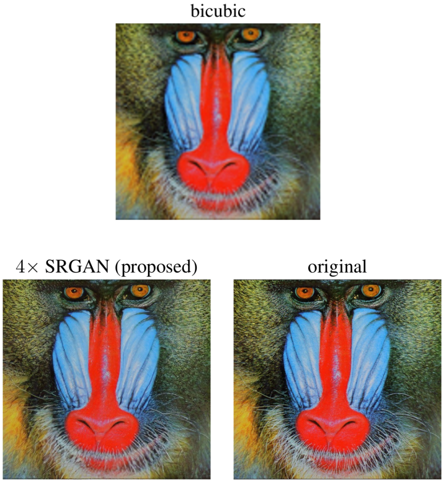


arXiv: 1609.04802

- Single-image super-resolution (SR): estimation of a high resolution (HR) image from a single low resolution (LR) image
- Intensively studied in the field of image processing
- Has been applied to jet physics (arXiv: 2012.11944)
- State-of-the-art models based on deep CNNs
  - Trained on LR-HR image pairs
- GAN-based training setups particularly successfull in producing realistic SR images

Photon calorimetry

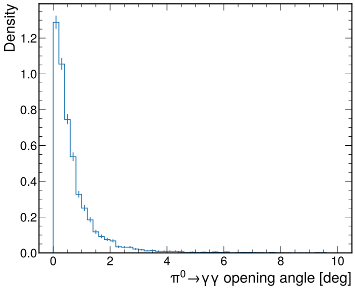
Can we improve photon reconstruction by learning from a better calorimeter?

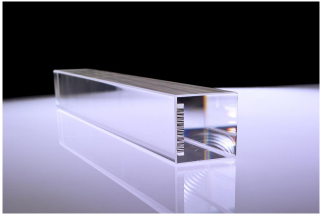


arXiv: 1609.04802

Simulated samples

- Geant 4 simulation of γ and $\pi^0 \to \gamma\gamma$ at 20 GeV
  - Remove $\pi^0 \to \gamma\gamma$ without strongly collimated photons (angle $> 2°$)
- Simplified $PbWO_4$ electromagnetic calorimeter
  - LR granularity adapted from CMS barrel ECAL
  - LR: 24×24 crystals, 2.2 cm $\times$ 2.2 cm $\times$ 23 cm each
  - HR: 96×96 crystals, 0.55 cm $\times$ 0.55 cm $\times$ 23 cm each
- Simulation of LR-HR image pairs
- Cut shower images to 6×6 (LR) and 24×24 (HR) crystals

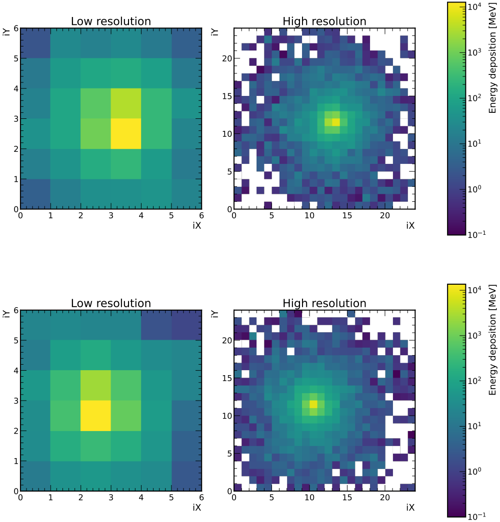




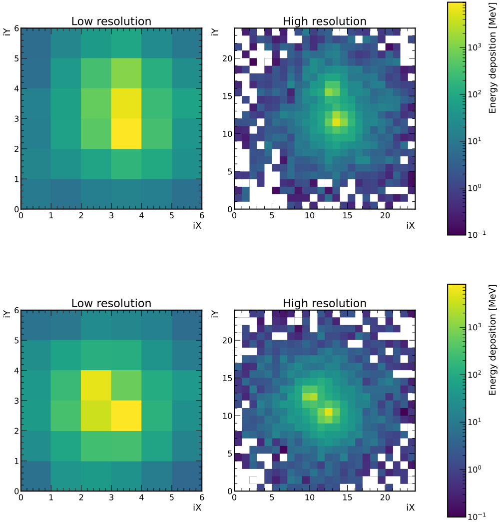
$PbWO_4$ crystal of CMS ECAL
http://cds.cern.ch/record/1101276

Simulated samples

- Example photons:

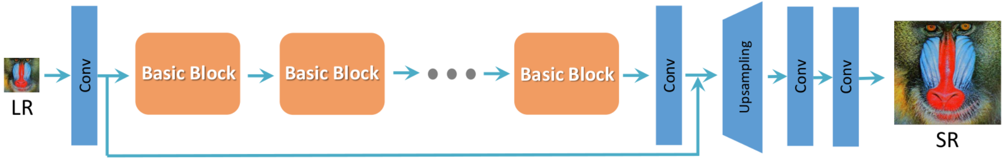


- Example pions:

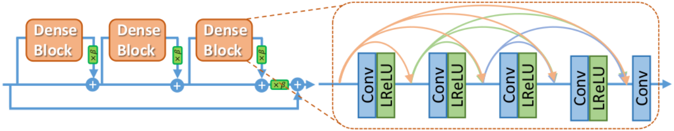


Model architecture

- Model architecture inspired by *Enhanced Super-Resolution GAN (ESRGAN)*, arXiv:1809.00219
- Generator consists of very deep CNN
- Most calculations done in LR feature space before upsampling

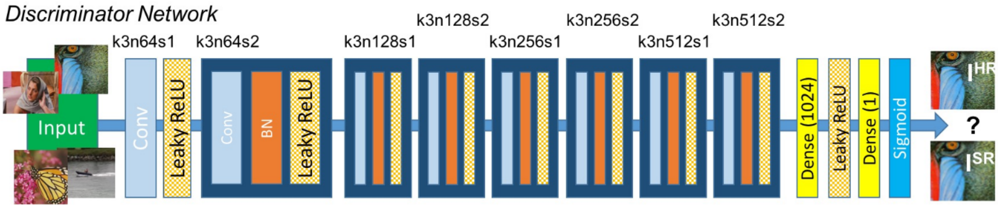


- Basic block: Residual-in-residual dense block (RRDB)



- We use 5 RRDBs, convolutional layers with 32 (3×3) kernels

- Discriminator inspired by SRGAN / ESRGAN



arXiv: 1609.04802

- Here: 6 convolutional layers, 2 dense layers
- Training strategy adapted to Wasserstein-GAN with gradient penalty (WGAN-GP, arXiv: 1704.00028)
  - Sigmoid output function removed
  - Batch Normalisation replaced by Layer Normalisation

- SRGAN is trained on 100k photon and 100k neutral pion examples
- As ESRGAN, we use the concept of perceptual loss (arXiv: 1603.08155)
  - Additional loss term for generator
  - Intead of a *per-pixel* loss, perceptual loss uses high-level features Φ extracted from pretrained CNNs

$$\mathcal{L}_{\mathrm{p}} \propto \left|\Phi(\mathrm{HR}) - \Phi(\mathrm{SR})\right|^2$$

  - ESRGAN: VGG-19 network (very deep CNN for image classification, arXiv: 1409.1556)
  - We use simple CNN trained on HR images to separate photons from pions

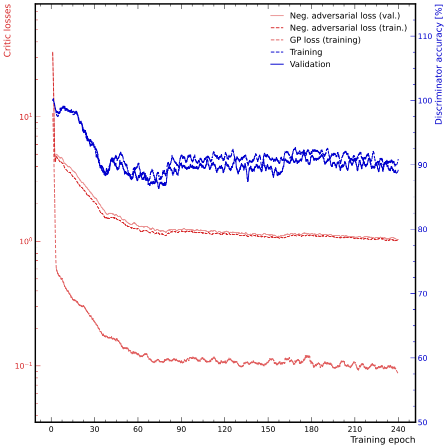
- Generator:

$$\mathcal{L}_{\text{gen}} = 10^{-4} \cdot \mathcal{L}_{\text{WGAN}} + 10^{-9} \cdot \mathcal{L}_{\text{perceptual}}$$

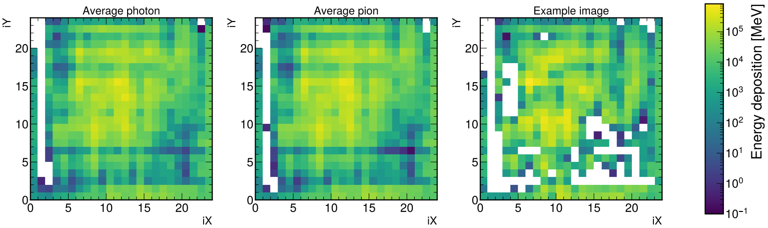
- Discriminator

$$\mathcal{L}_{\text{dis}} = 10 \cdot \mathcal{L}_{\text{WGAN}} + 10 \cdot \mathcal{L}_{\text{GP}}$$

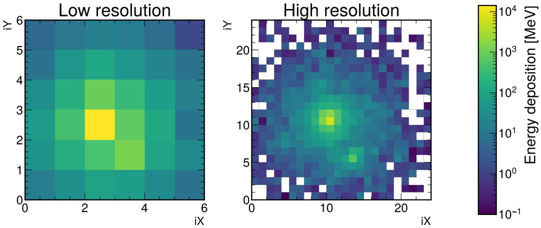
- Monitoring of physics-metrics during training:
  - Width of the shower
  - Pion rejection: how well can a CNN trained on HR images separate photons and pions?

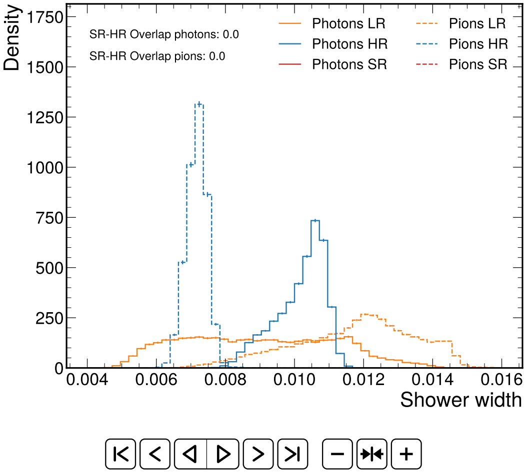


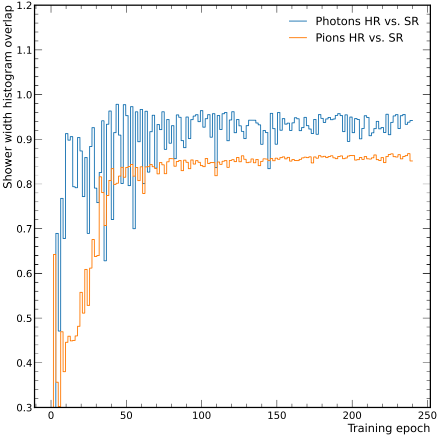
Example SR images (0 steps)



Example image

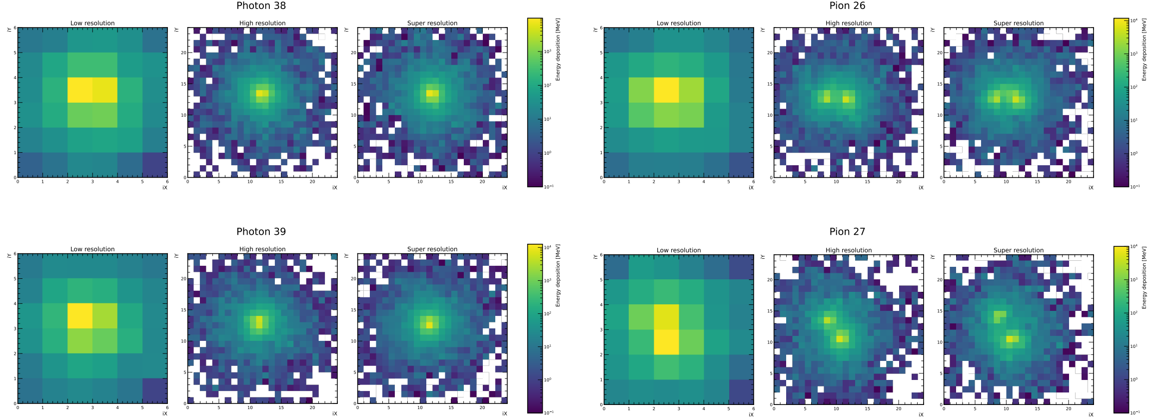






- Generator network implicitly learns to treat the two classes separately

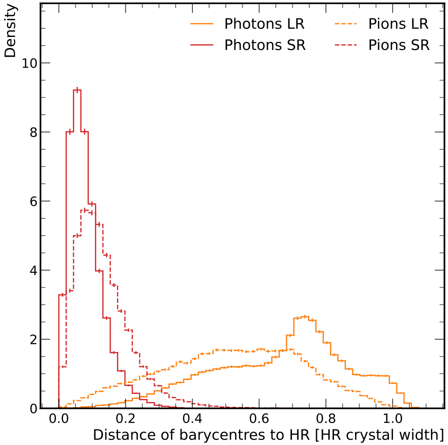
Example SR images



- SR images are very close to HR truth

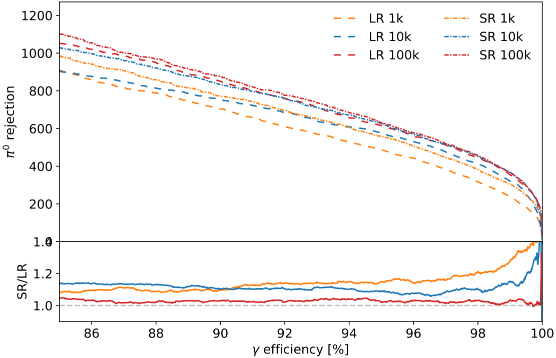
What is it good for?

- Barycentres of SR images are closer to truth than barycentres of LR images
- Improvement in order of one HR crystal width ($\approx 0.5\,\mathrm{cm}$)
- Improved spatial resolution $\rightarrow$ e.g. better mass resolution for diphoton events



What is it good for?

- SR as image preprocessing can help for training classifiers
- Here: simple CNNs with same depth and width for LR and SR
- Improved pion rejection over LR given limited training statistics
- In LHC experiments: typically low training statistics since only very small fraction of simulated jets passes photon preselection



- *Work in progress! Here: preliminary studies from previous detector simulation, less close to LHC experiment conditions*

Conclusions & outlook

- Super-resolution applied to photon calorimeter images
- Generator incorporates knowledge about shower development from HR calorimeter
- SR images are realistic estimations how a shower might look in a better calorimeter
- Spatial resolution is improved in SR
- Classifiers learn faster from SR images than from LR images

- *Ongoing work... Stay tuned!*

BACKUP

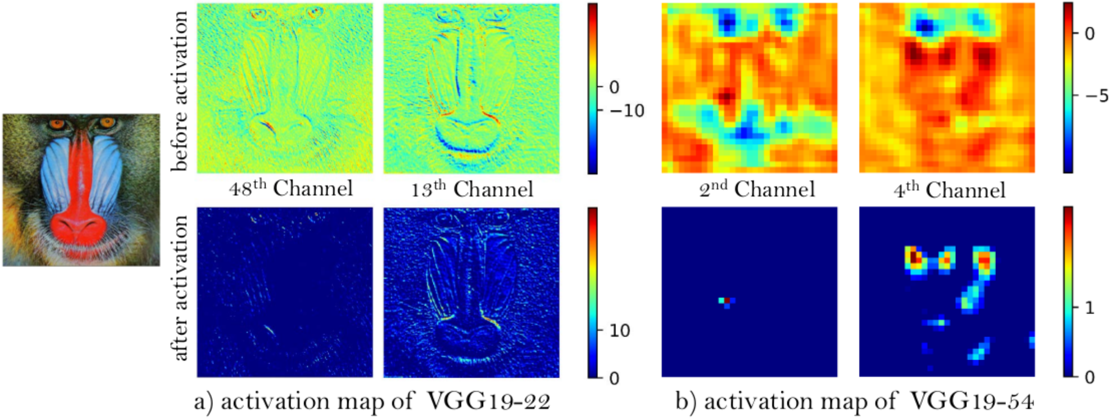
WGAN-GP algorithm

Algorithm 1 WGAN with gradient penalty. We use default values of $\lambda = 10$, $n_{\text{critic}} = 5$, $\alpha = 0.0001$, $\beta_1 = 0$, $\beta_2 = 0.9$.

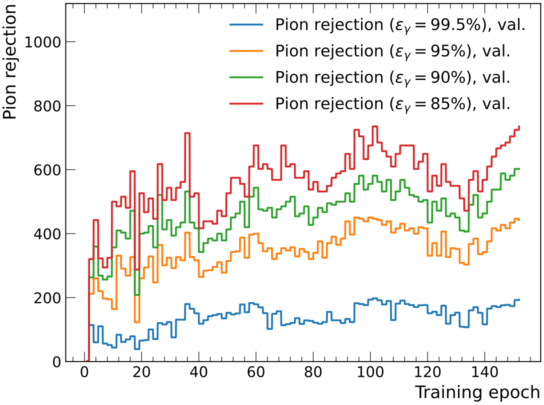
Require: The gradient penalty coefficient λ, the number of critic iterations per generator iteration n_{critic}, the batch size m, Adam hyperparameters α, β_1, β_2.
Require: initial critic parameters w_0, initial generator parameters θ_0.

while θ has not converged **do**
    for $t = 1, \ldots, n_{\text{critic}}$ **do**
        for $i = 1, \ldots, m$ **do**
            Sample real data $\boldsymbol{x} \sim \mathbb{P}_r$, latent variable $\boldsymbol{z} \sim p(\boldsymbol{z})$, a random number $\epsilon \sim U[0,1]$.
            $\tilde{\boldsymbol{x}} \leftarrow G_\theta(\boldsymbol{z})$
            $\hat{\boldsymbol{x}} \leftarrow \epsilon \boldsymbol{x} + (1-\epsilon)\tilde{\boldsymbol{x}}$
            $L^{(i)} \leftarrow D_w(\tilde{\boldsymbol{x}}) - D_w(\boldsymbol{x}) + \lambda(\|\nabla_{\hat{\boldsymbol{x}}} D_w(\hat{\boldsymbol{x}})\|_2 - 1)^2$
        end for
        $w \leftarrow \text{Adam}(\nabla_w \frac{1}{m}\sum_{i=1}^{m} L^{(i)}, w, \alpha, \beta_1, \beta_2)$
    end for
    Sample a batch of latent variables $\{\boldsymbol{z}^{(i)}\}_{i=1}^{m} \sim p(\boldsymbol{z})$.
    $\theta \leftarrow \text{Adam}(\nabla_\theta \frac{1}{m}\sum_{i=1}^{m} -D_w(G_\theta(\boldsymbol{z})), \theta, \alpha, \beta_1, \beta_2)$
end while

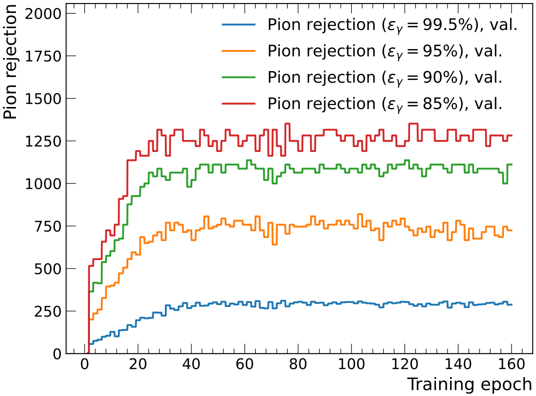
Perceptual loss ESRGAN



Perceptual loss: training



Without perceptual loss



With perceptual loss