ML4Jets2020, 01/17/2020, NYU

The Di-Higgs Photography with Deep Neural Networks

Jeong Han Kim

UNIVERSITY OF NOTRE DAME

With Minho Kim, K. C. Kong, Konstantin T. Matchev, Myeonghun Park

- [Phys.Rev.Lett. 122 (2019) no.9, 091801]
- [JHEP 1909 (2019) 047]
- [arXiv: In Progress]

Motivation

- After a long period of direct searches in the LHC, no tantalizing evidence of new physics has been found.

- Among many other possibilities, if the scale of new physics is too heavy to reach:

- The EFT provides a framework to encode new physics effects in higher dimensional operators.

$$\mathcal{L} = \mathcal{L}_{SM} + \frac{c_1^{(6)}}{\Lambda_{NP}^2}\mathcal{O}_1^{(6)} + \frac{c_2^{(6)}}{\Lambda_{NP}^2}\mathcal{O}_2^{(6)} + \cdots$$

$\Lambda_{\rm EW}$

?

Λ_{NP}

A Bridge toward BSM

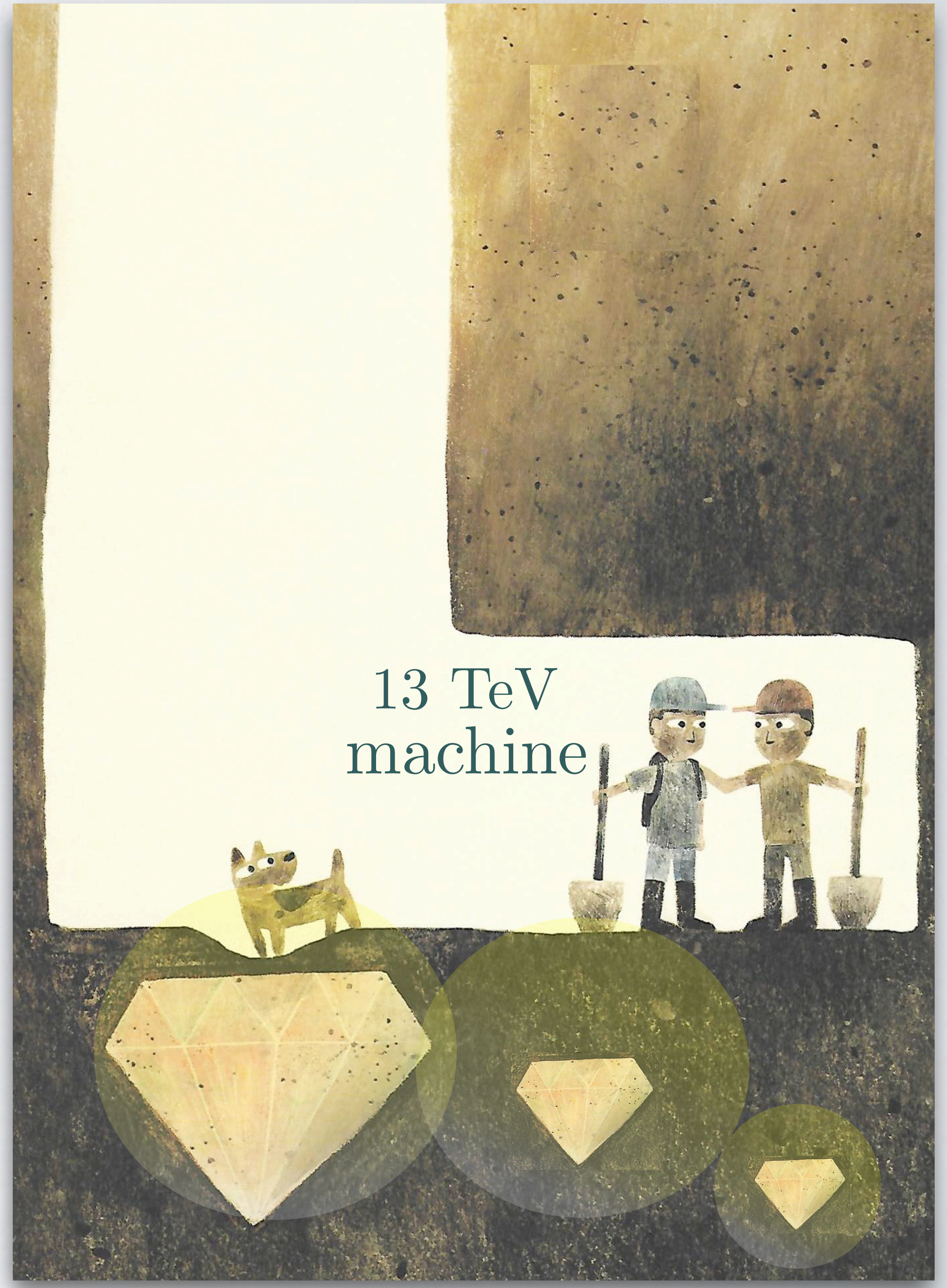
- Precision measurements of these coefficients will tell us two things:
- What would be a rough scale of new physics.
- What would be a possible structure of the underlying new physics.

$$\mathcal{L} = \mathcal{L}_{SM} + \frac{c_1^{(6)}}{\Lambda_{NP}^2}\mathcal{O}_1^{(6)} + \frac{c_2^{(6)}}{\Lambda_{NP}^2}\mathcal{O}_2^{(6)} + \cdots$$

$\Lambda_{\rm EW}$

?

Λ_{NP}



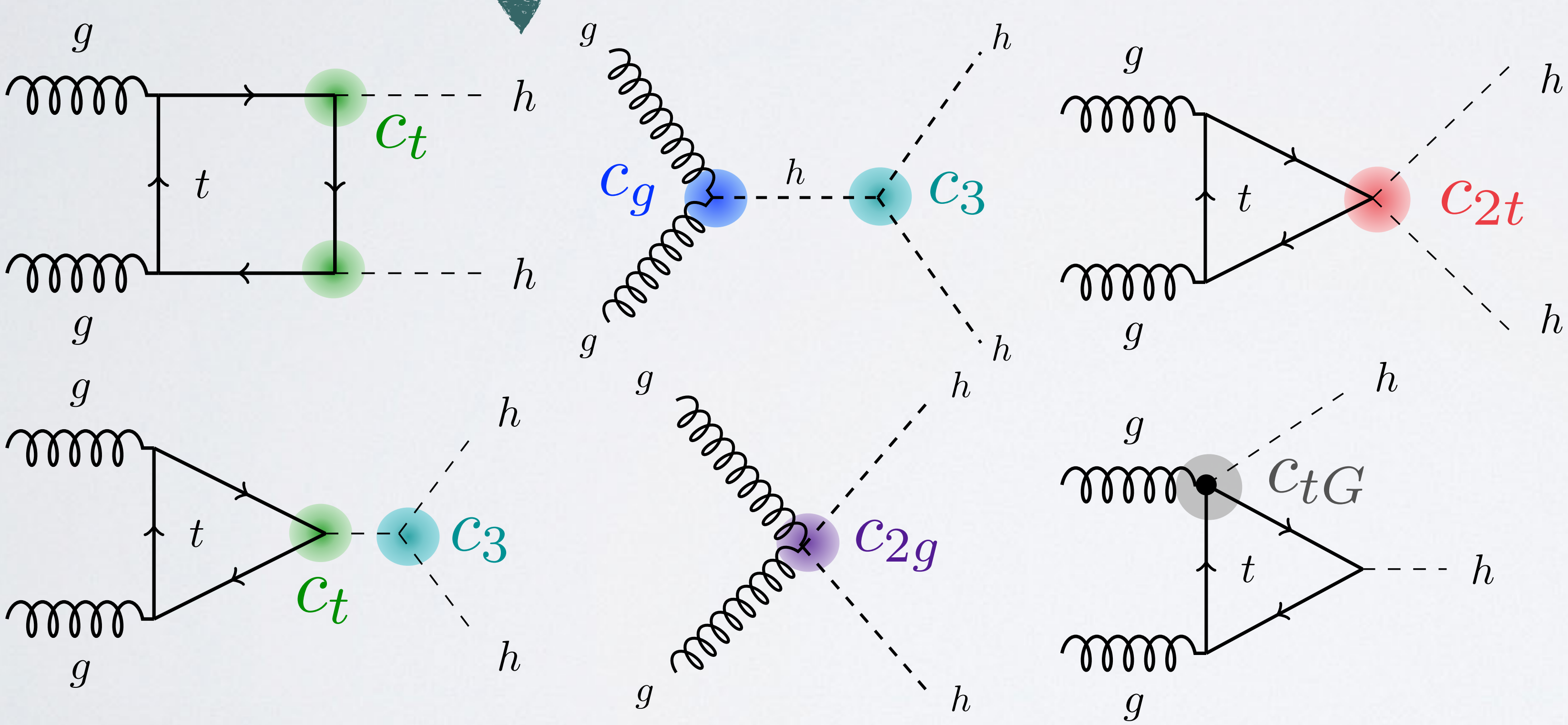
Double Higgs as a Benchmark Process

$$\mathcal{L} = \mathcal{L}_{SM} + \boxed{\Delta\mathcal{L}_6} + \cdots$$

$$+\frac{\bar{c}_{tG}\, g_S}{m_W^2}(\overline{Q}_L H^c)\sigma^{\mu\nu}T^a t_R G^a_{\mu\nu} + \text{h.c.}$$

$$\frac{\bar{c}_H}{2v^2}(\partial^\mu |H|^2)^2 - \frac{\bar{c}_6}{v^2}\left(\frac{m_h^2}{2v^2}\right)|H|^6 + \frac{\bar{c}_u}{v^2} y_t(|H|^2 \overline{Q}_L H^c t_R + \text{h.c.}) + \bar{c}_g\left(\frac{g_s^2}{m_w^2}\right)|H|^2 G^a_{\mu\nu}G^{a\,\mu\nu}$$

After EWSB…



- We start from the *hh* as a benchmark final state where there are five dimension 6-operators that directly contribute to the production.
- They give rise to anomalous Higgs couplings.
- The *hh* can be a main playground to search for the hint of new physics.
- We are mainly interested in the c_3 to probe the underlying principle of EWSB.

Discovery potential of the SM *hh* at HL-LHC ?

	Channels	Significances	Combined	Combined
ATLAS ATL-PHYS-PUB-2018-053	$bb\gamma\gamma$	2.1	3.5	$\boxed{\mathbf{4.5}}$
	$bb\tau\tau$ *(fully-hadronic)* + $bb\tau\tau$ *(semi-leptonic)*	2.5		
	$bbbb$	1.4		
CMS CMS-FTR-18-019-PAS	$bb\gamma\gamma$	1.8	2.8	
	$bb\tau\tau$ *(fully-hadronic)* + $bb\tau\tau$ *(semi-leptonic)*	1.6		
	$bbbb$	1.2		
	$bbWW^*$ ($\ell\ell\nu\nu$)	0.59		
	$bbZZ^*$ ($\ell\ell\ell\ell$)	0.37		

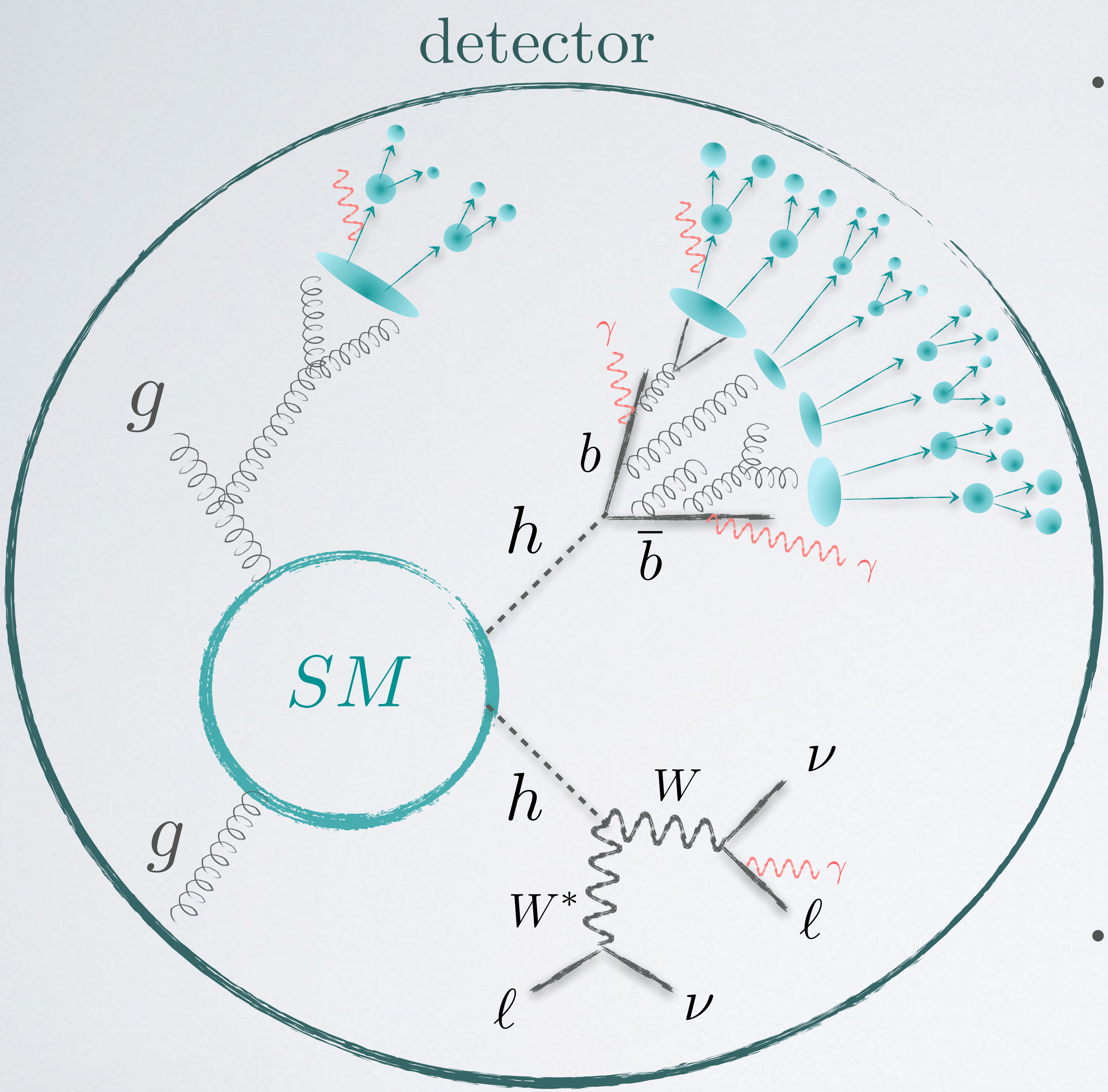
$3\ \mathrm{ab}^{-1}(14\ \mathrm{TeV})$

$c_3 = 1$

Previously overlooked and less studied

- In order to access analomous Higgs couplings we should observe the *hh* first.
- Since no single channel gives us 5σ , combining channels is essential.
- Recent analyses show that the combined significance is expected to be 4.5σ at the HL-LHC.
- Any potential improvement on the individual channel means a lot to discover the *hh*.

Previously on $hh \rightarrow bbWW^*$



- $hh \rightarrow bbWW^*$ channel suffers for the large $t\bar{t}$ background where the final state is identical. CMS-FTR-15-002-PAS

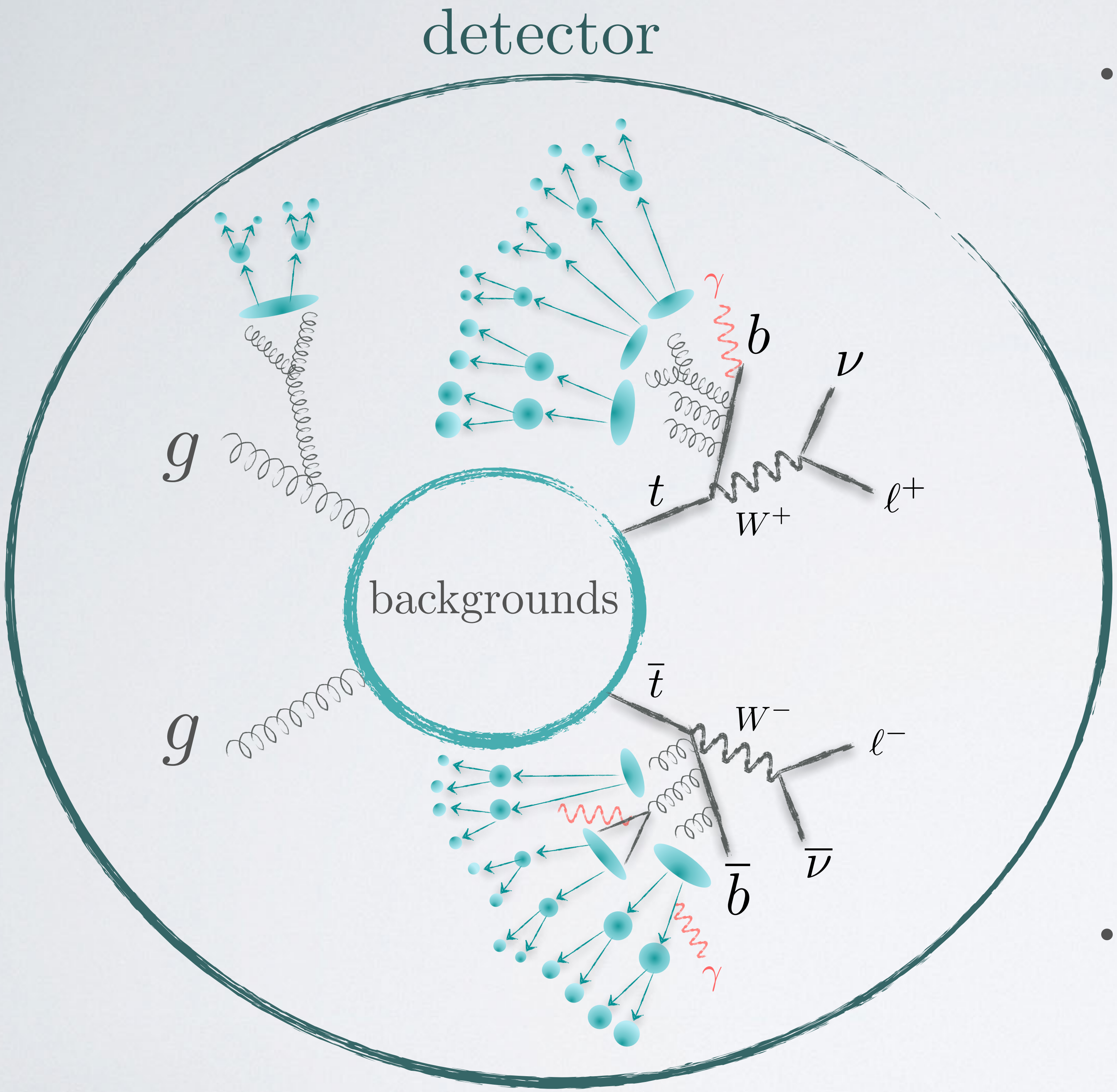
  Dolan, Englert, Spannowsky [2012]

  Adhikary, Banerjee, Barman, Bhattacherjee, Niyogi [2017]

- A challenging task for a traditional shape analysis which utilizes only a subset of information (e.g. p_T, m_{jj}, ΔR_{jj} , …)

- A neural network is an ideal technique to disentangle the hh and $t\bar{t}$.

Previously on $hh \to bbWW^*$



- $hh \to bbWW^*$ channel suffers for the large $t\bar{t}$ background where the final state is identical. CMS-FTR-15-002-PAS

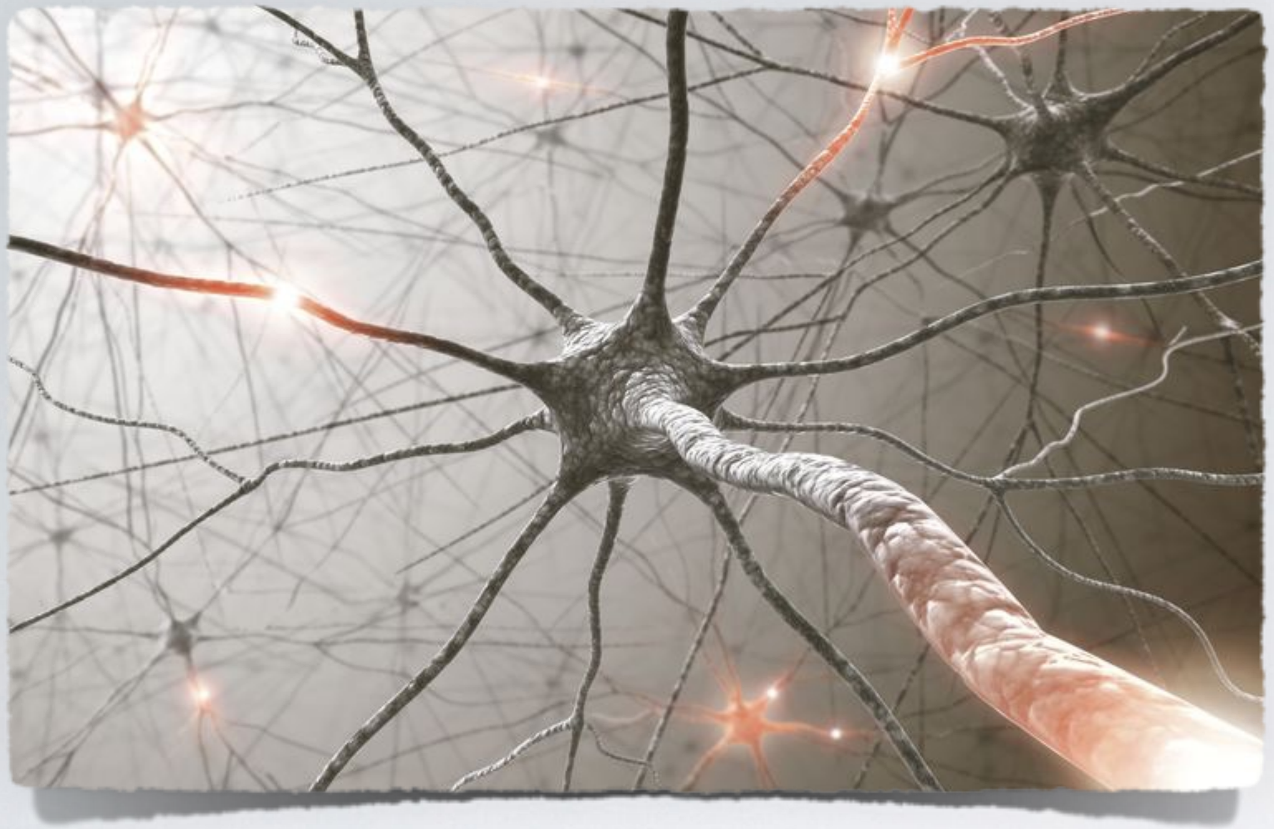
Dolan, Englert, Spannowsky [2012]

Adhikary, Banerjee, Barman, Bhattacherjee, Niyogi [2017]

- A challenging task for a traditional shape analysis which utilizes only a subset of information (e.g. p_T, m_{jj}, ΔR_{jj} , …)

- A neural network is an ideal technique to disentangle the hh and $t\bar{t}$.

Deep Neural Network



- Inputs to our neural network are hadrons, lepton, photon, and reconstructed neutrino images. J. Gallicchio and M. D. Schwartz [2010]

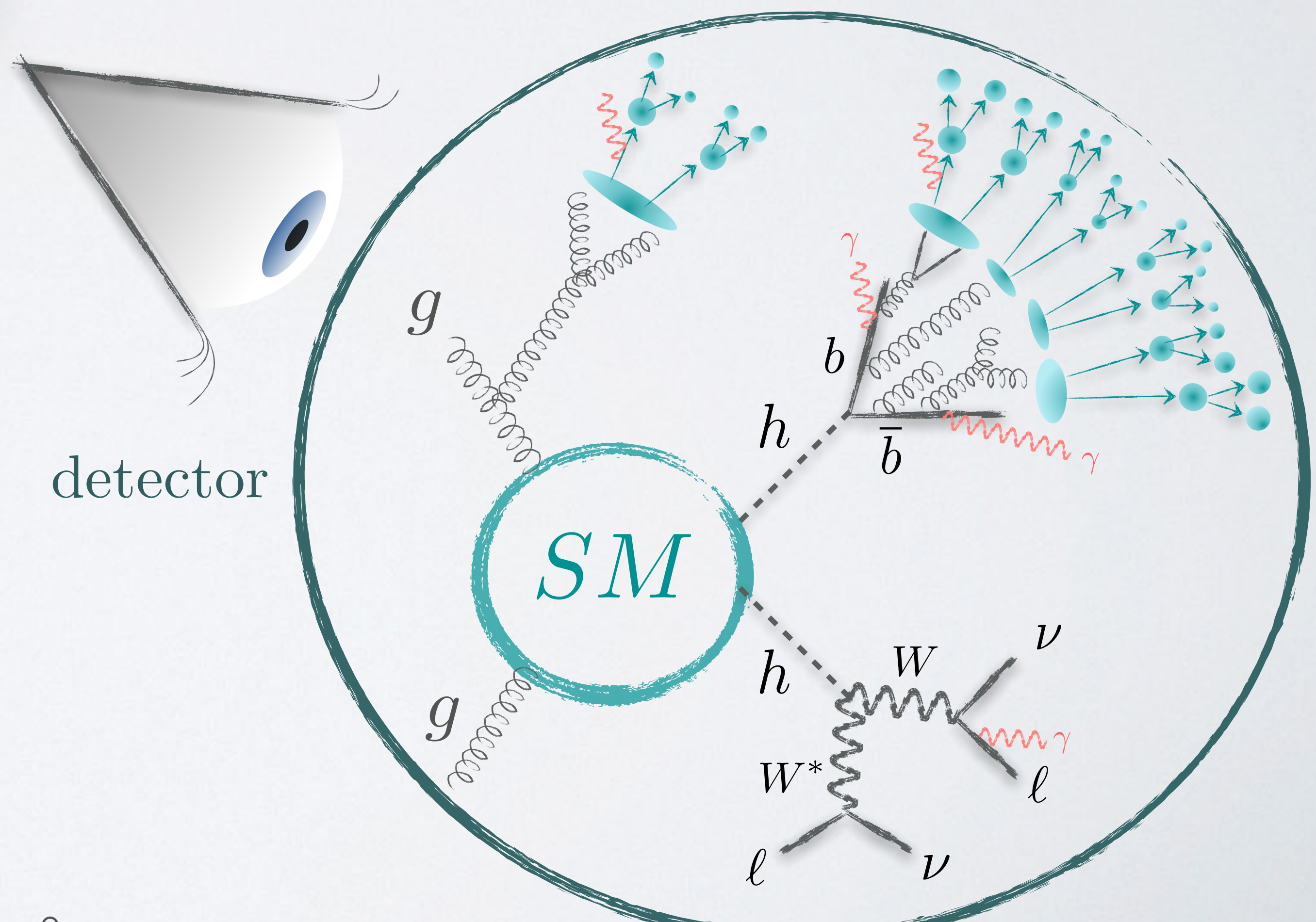
L. Oliveira, M. Kagan, L. Machkey, B. Nachman, A. Schwartzman [2015]
M. Farina, Y. Nakai, D. Shih [2018]
S. Diefenbacher, H. Frost, G. Kasieczka, T. Plehn, J. M. Thompson [2019]
See also P. T. Komiske, E. M. Metodiev, J. Thaler [2019]
A. Chakraborty, S. Lim, M. Nojiri [2019]
J. H. Kim, M. Kim, K. C. Kong, K. T. Matchev, M. Park [2019]
…

- I guess the method is already familiar, and used in many different context…

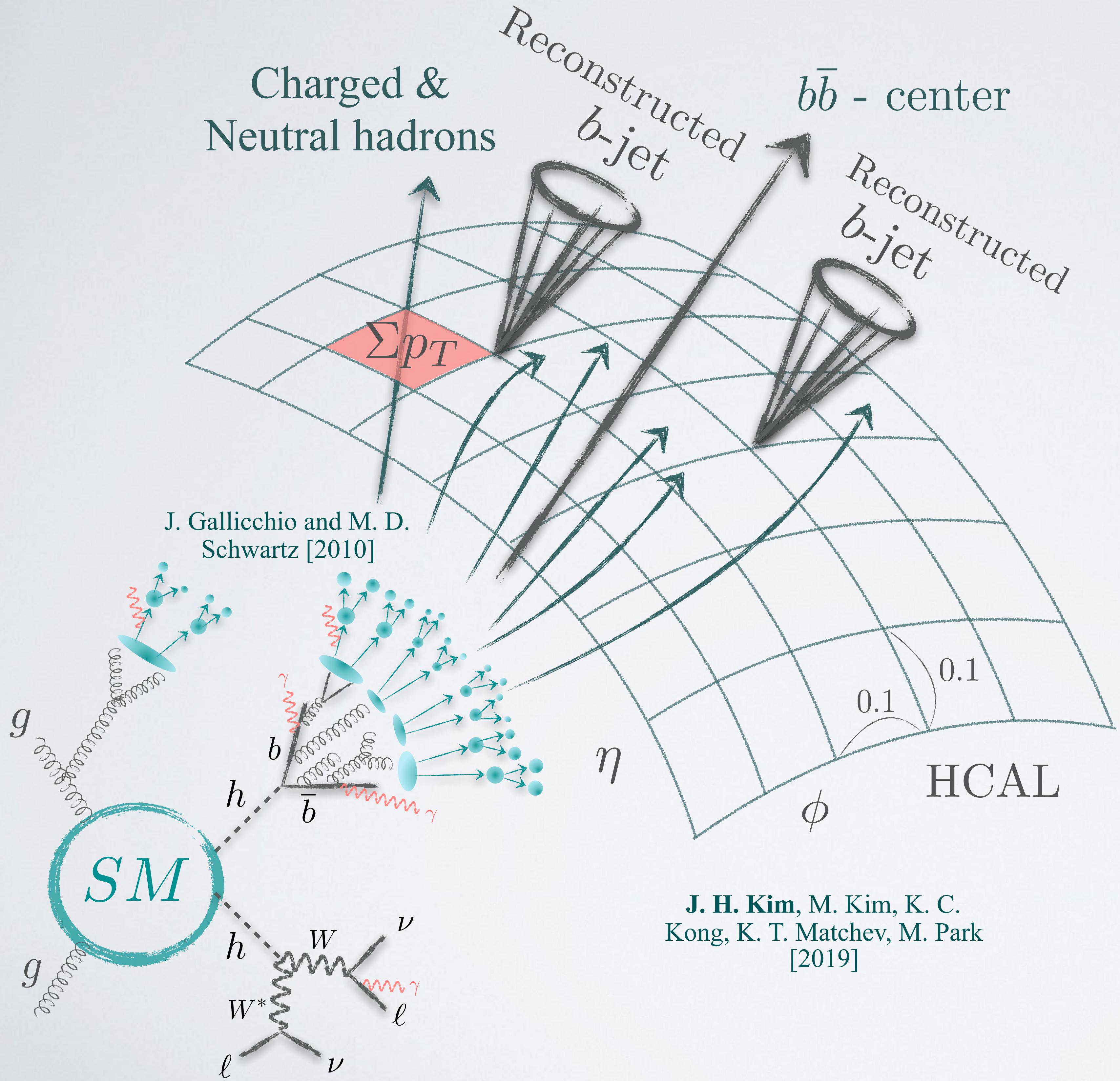
- The punchline is that we visualize the full final state as a set of images.

- We employ a deep neural network to learn a generic ME information.

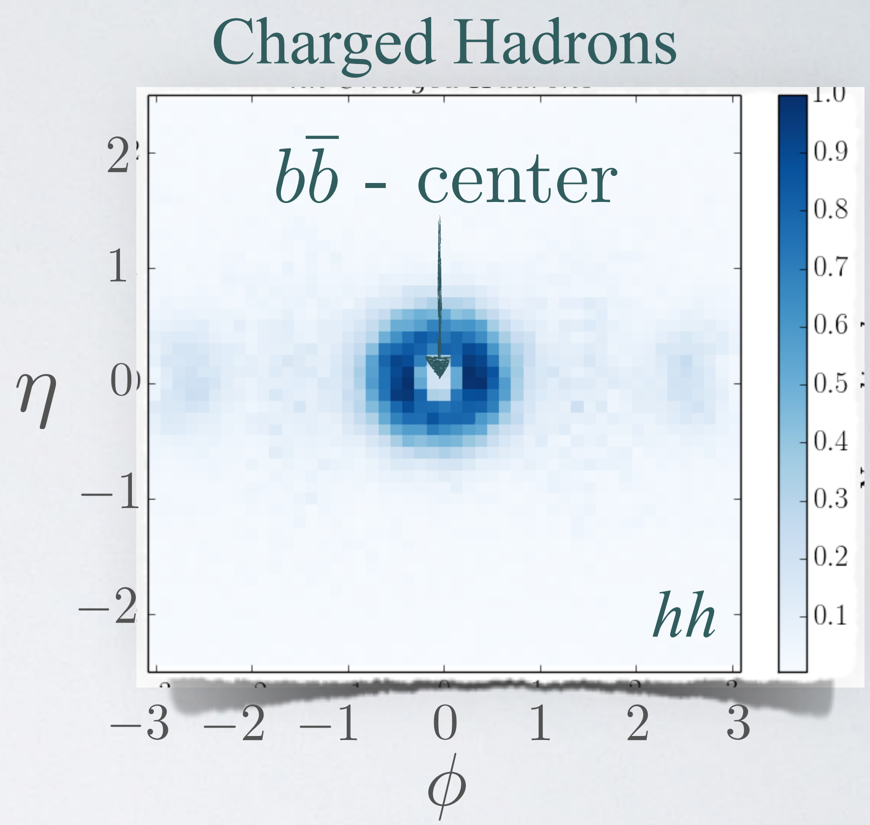
- As a pilot study, we will assume the SM couplings, and focus on disentangling the *hh* against the $t\bar{t}$ background.

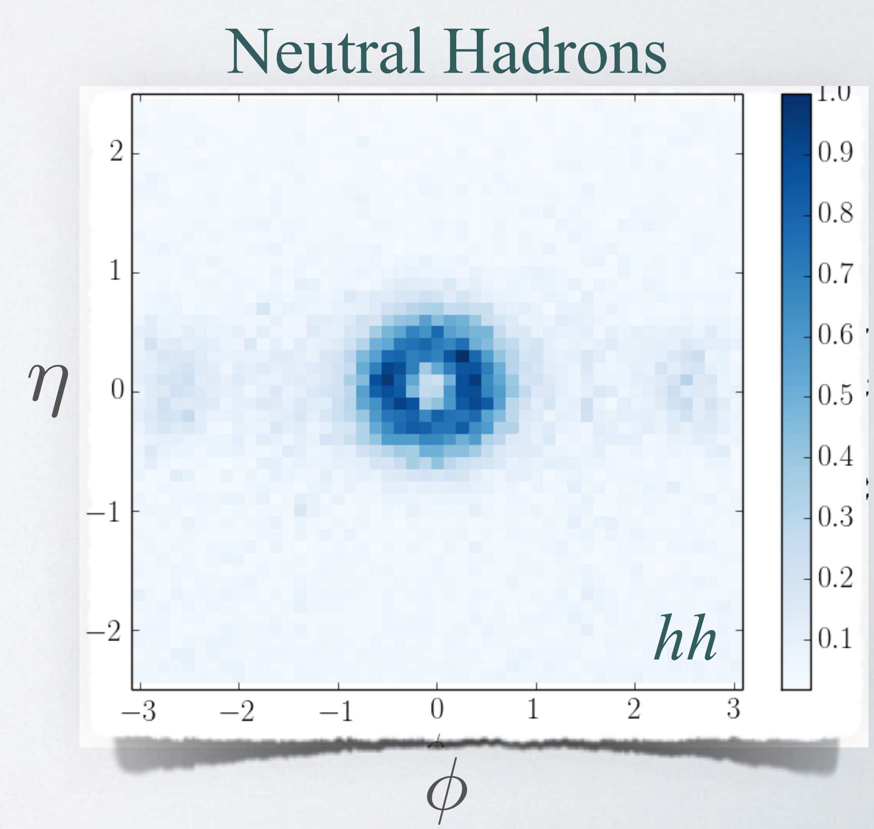


Processing Hadron Images



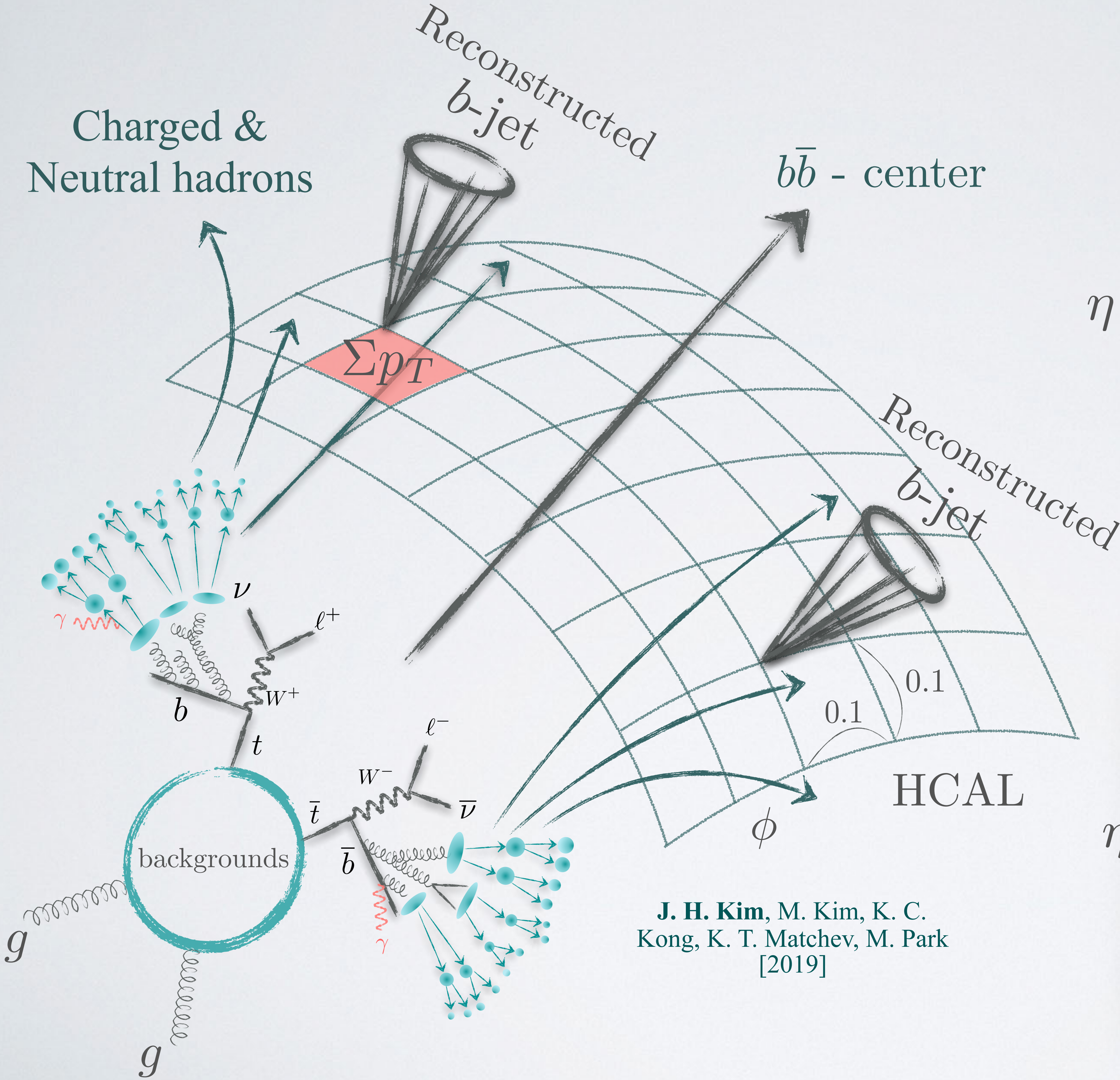
J. Gallicchio and M. D. Schwartz [2010]





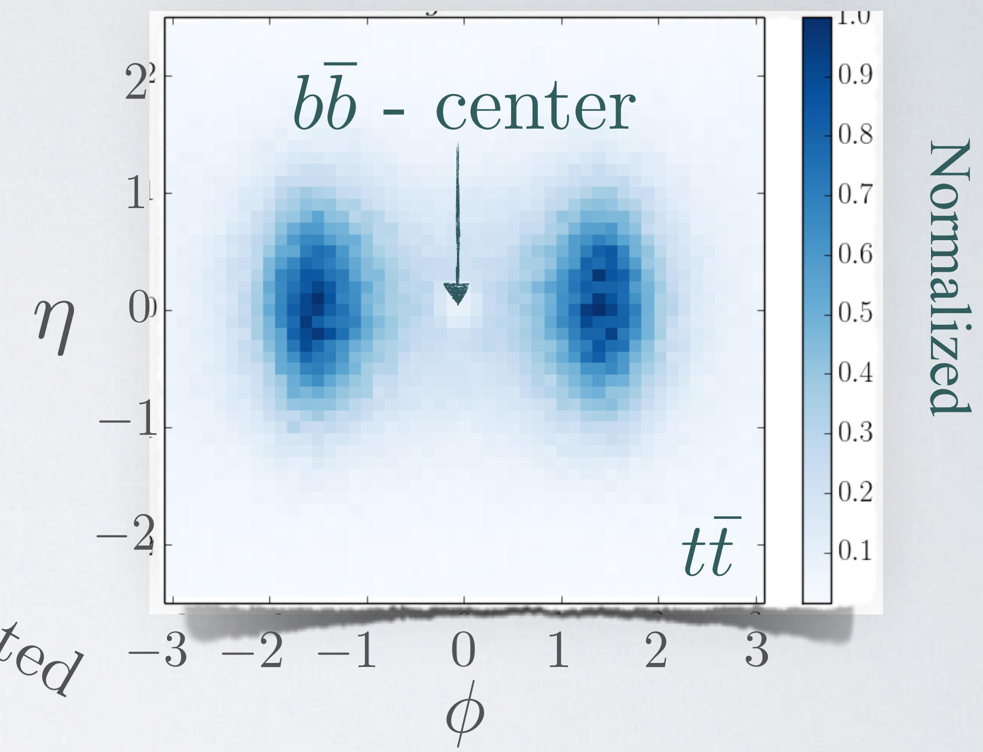
J. H. Kim, M. Kim, K. C. Kong, K. T. Matchev, M. Park [2019]

Processing Hadron Images

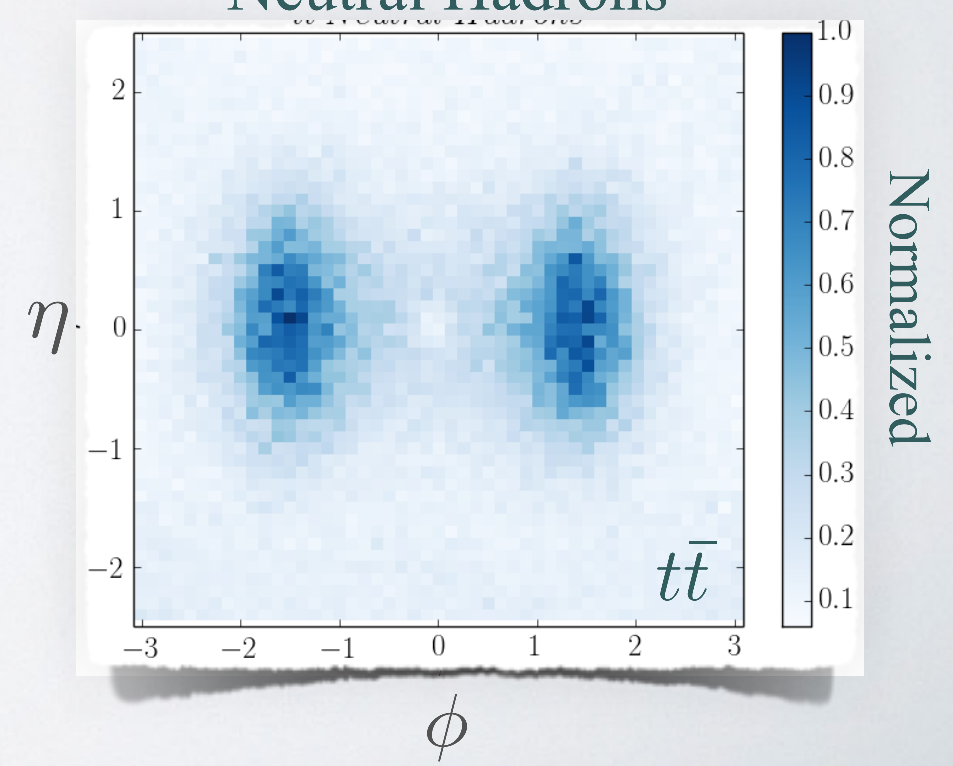


J. H. Kim, M. Kim, K. C. Kong, K. T. Matchev, M. Park [2019]

Charged Hadrons

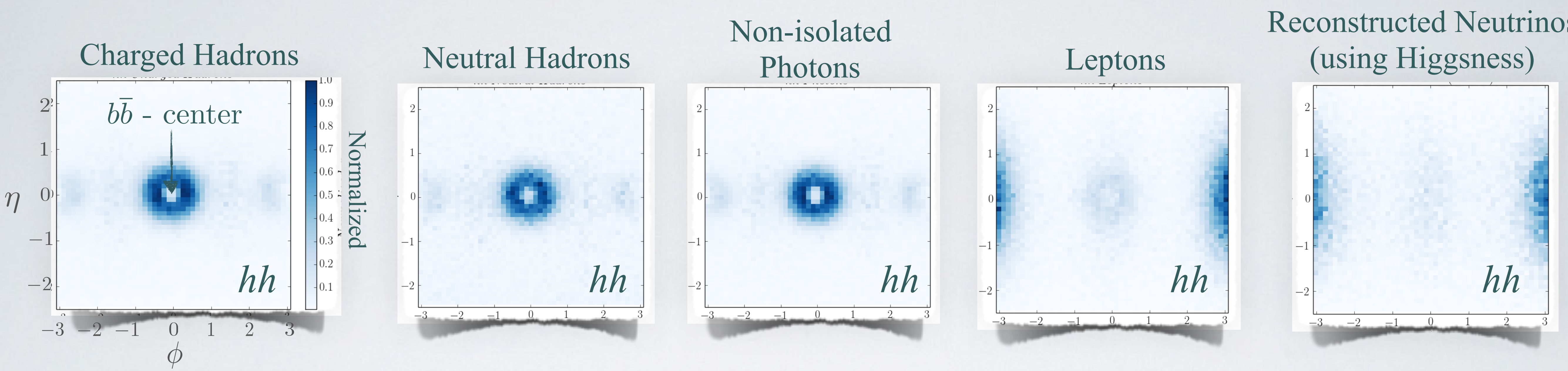


Neutral Hadrons

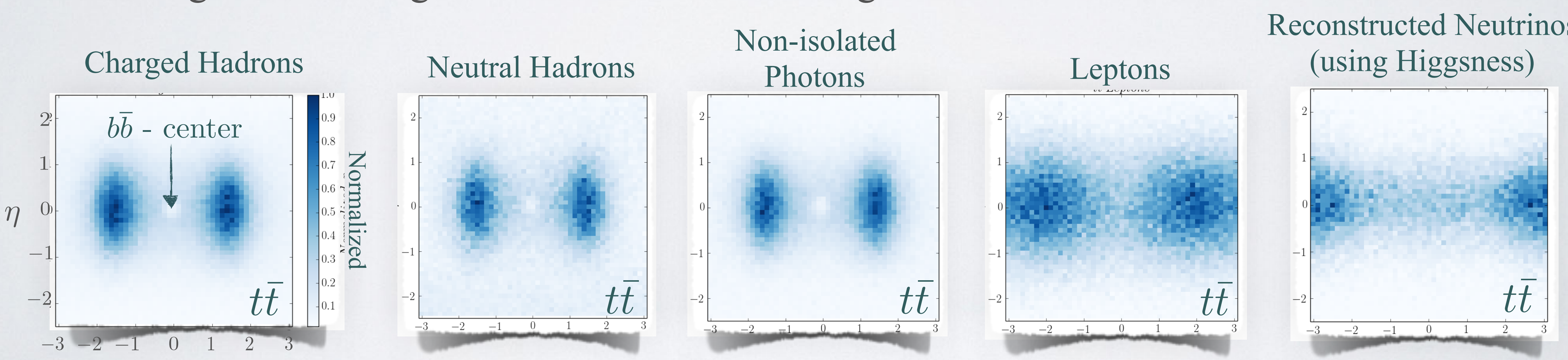


The Di-Higgs Photography

- Totally hadrons, lepton, photon, and neutrino images are shown for *hh*.



- Analogous five images are shown for $t\bar{t}$ background.



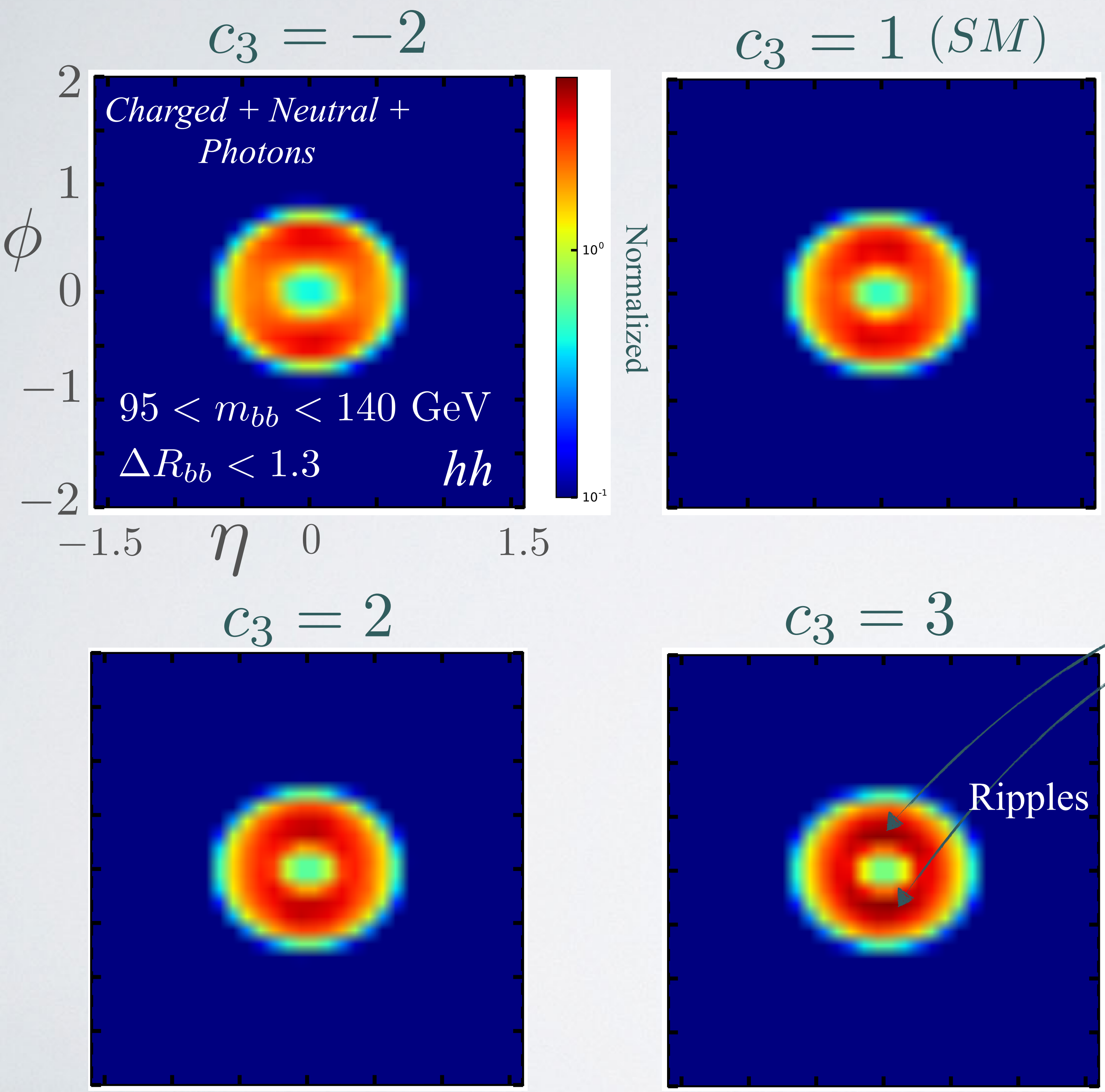
- Four-momenta of neutrinos are reconstructed using Higgsness and Topness, which implicitly utilize the mass information, m_h, m_t, m_W, m_{W^*}.

J. H. Kim, K. C. Kong, K. T. Matchev, M. Park [2018]

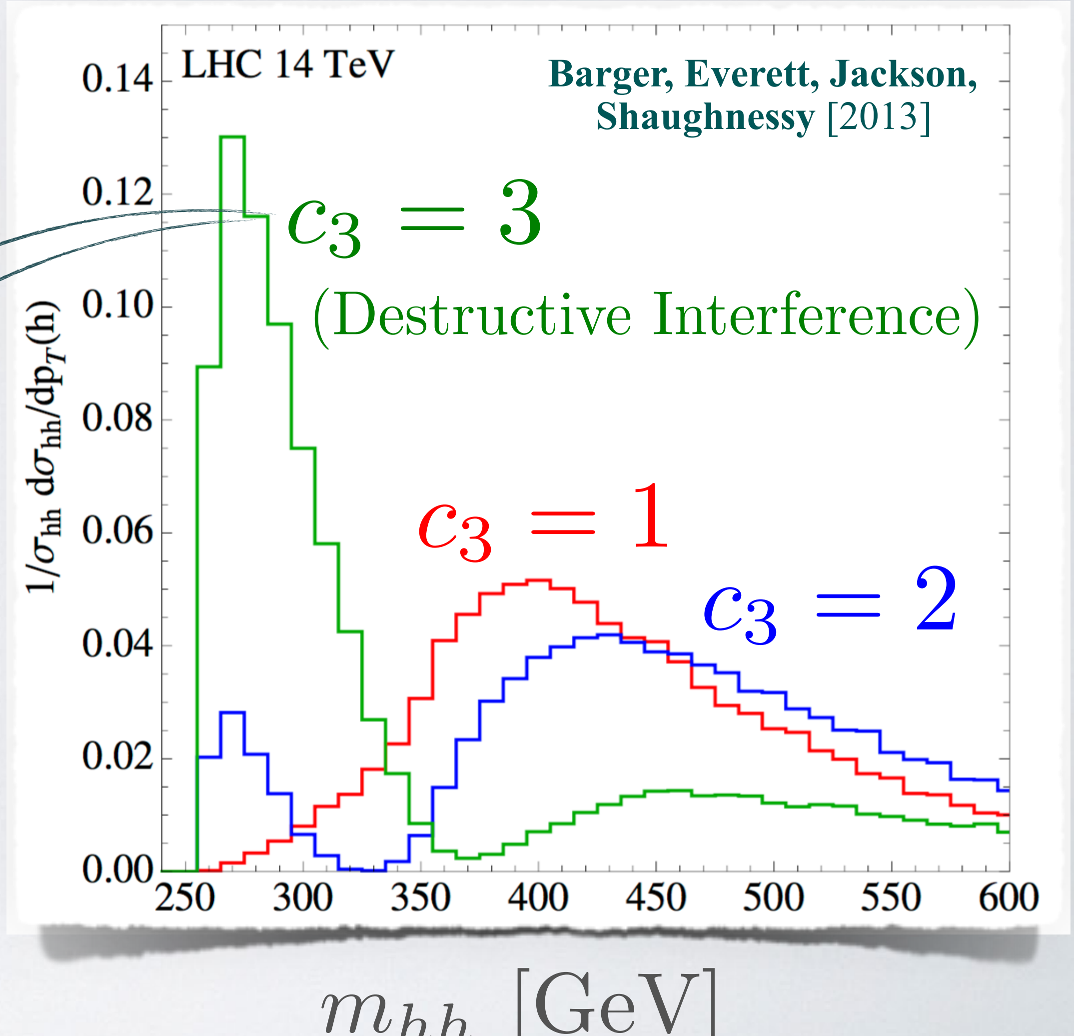
Shifting the Higgs triple coupling c_3

- How does the *hh* image vary by shifting the Higgs triple coupling c_3 ?

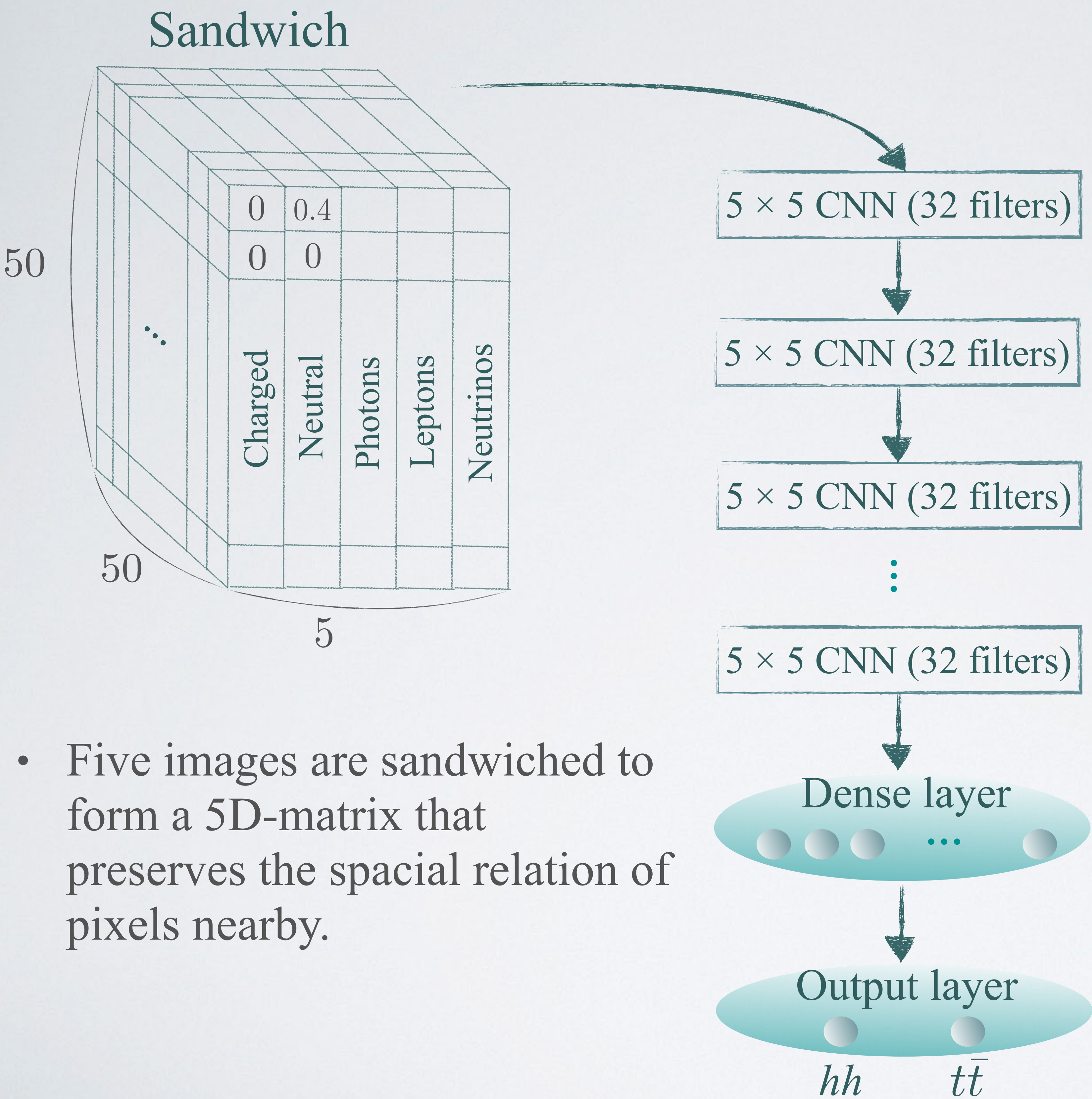
$$\mathcal{L} \supset -c_3 \frac{m_h^2}{2v} h^3$$



- Interestingly, the gradient of images change as moving into the region of a destructive interference.
- If the neural network is sensitive enough to pick up this small feature, that would be great.



Traditional Convolutional Neural Networks (CNN)



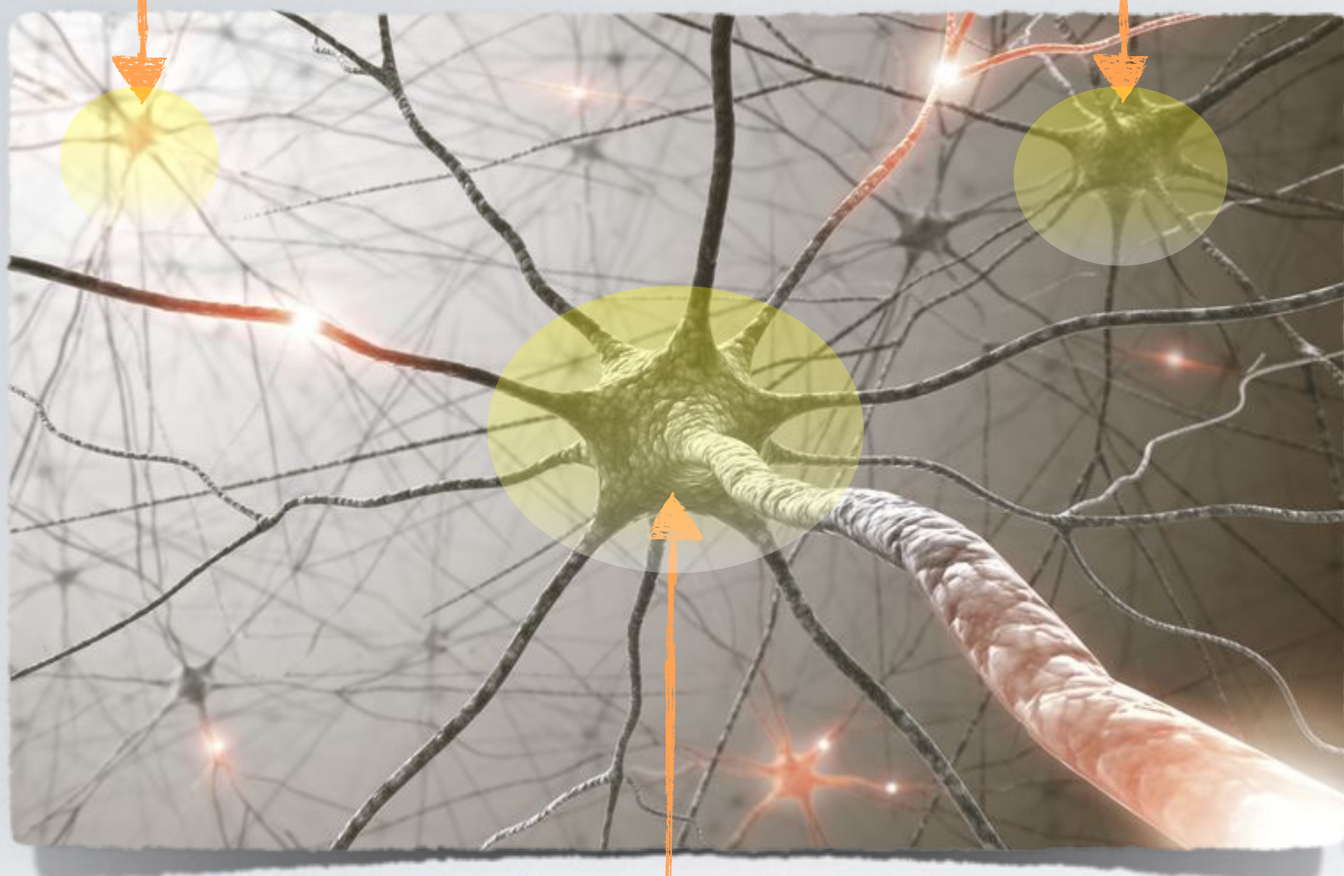
- Five images are sandwiched to form a 5D-matrix that preserves the spacial relation of pixels nearby.

- That is an input to a convolutional neural network (CNN).
- It is customary to stack a multiple CNN.
- Finally put the dense layer to reorganize the network.

A Guiding Principle of CNN

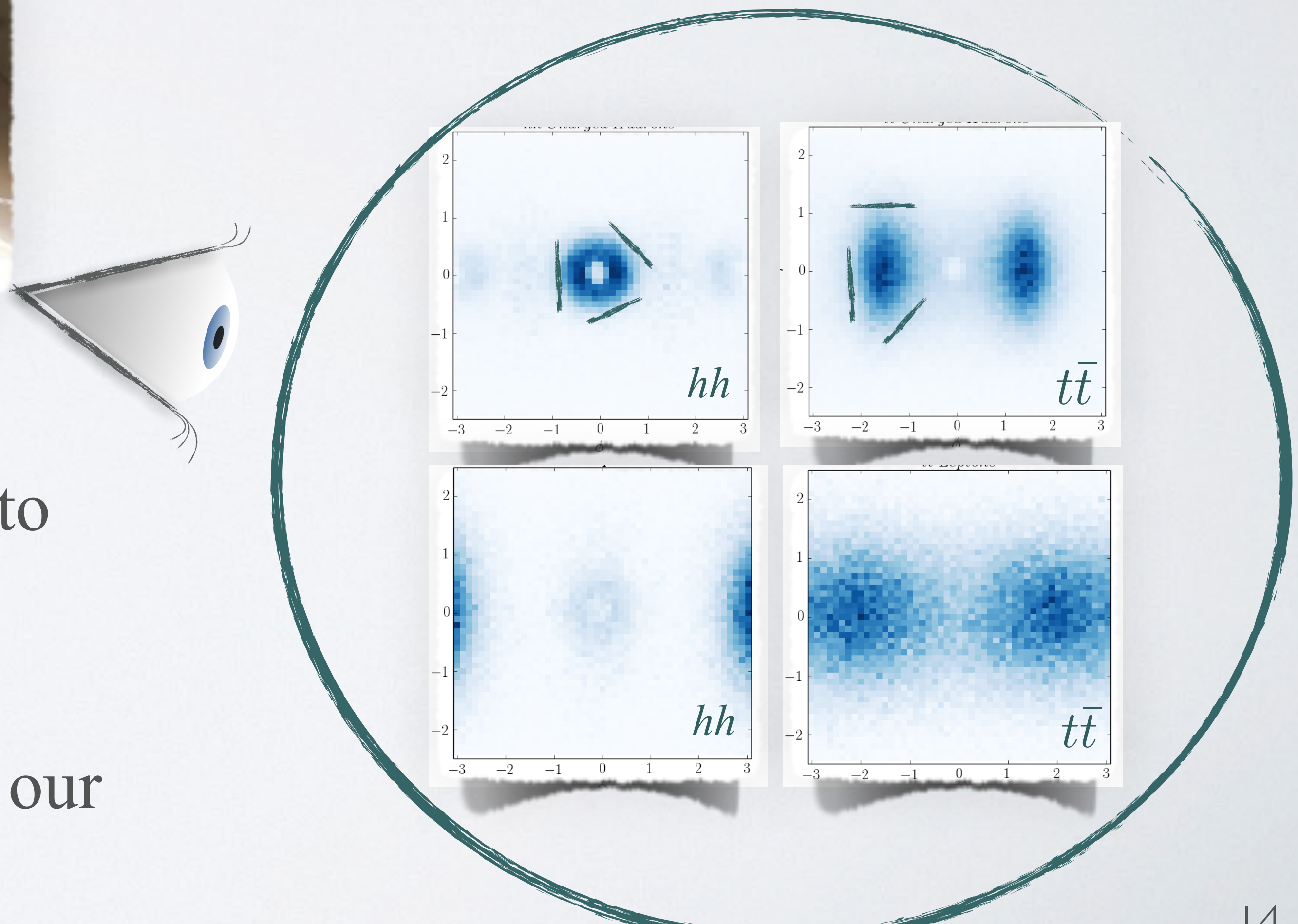
Sensitive to full objects

Sensitive to basic shapes



Sensitive to edges and bars

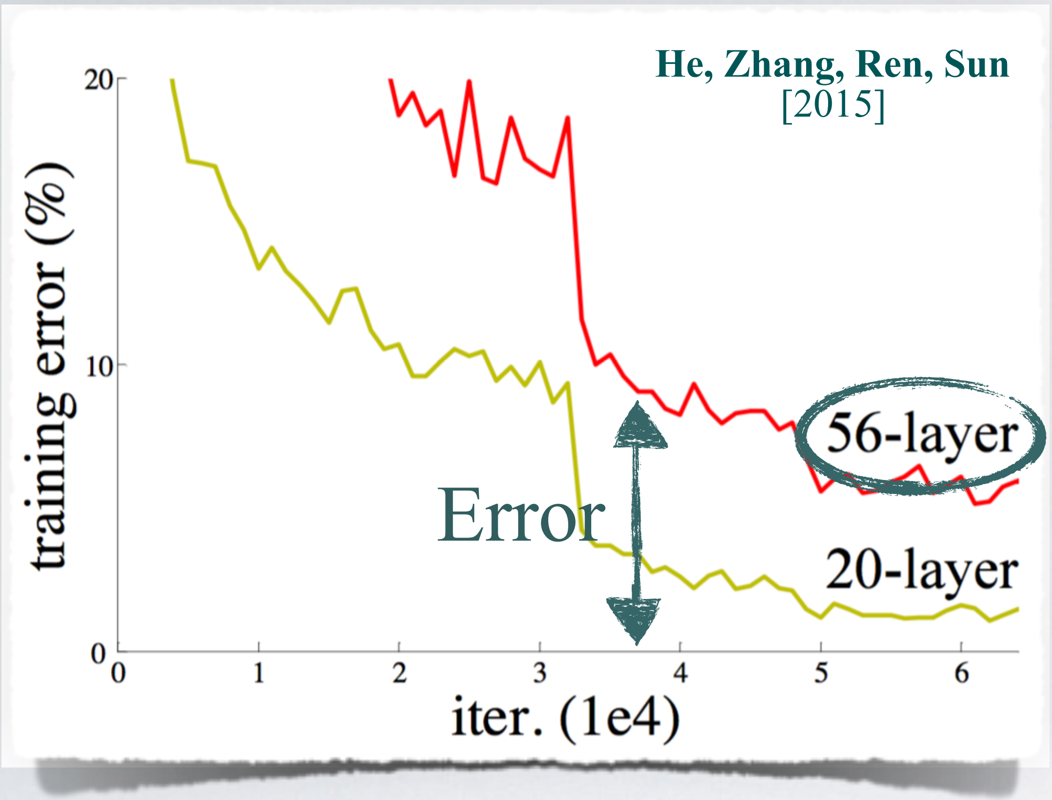
- There is a common trend that a network architecture needs to go deeper.
- This follows a lesson from the neurology as a guiding principle.
- Shallower neurons are only sensitive to edges and bars of the images.

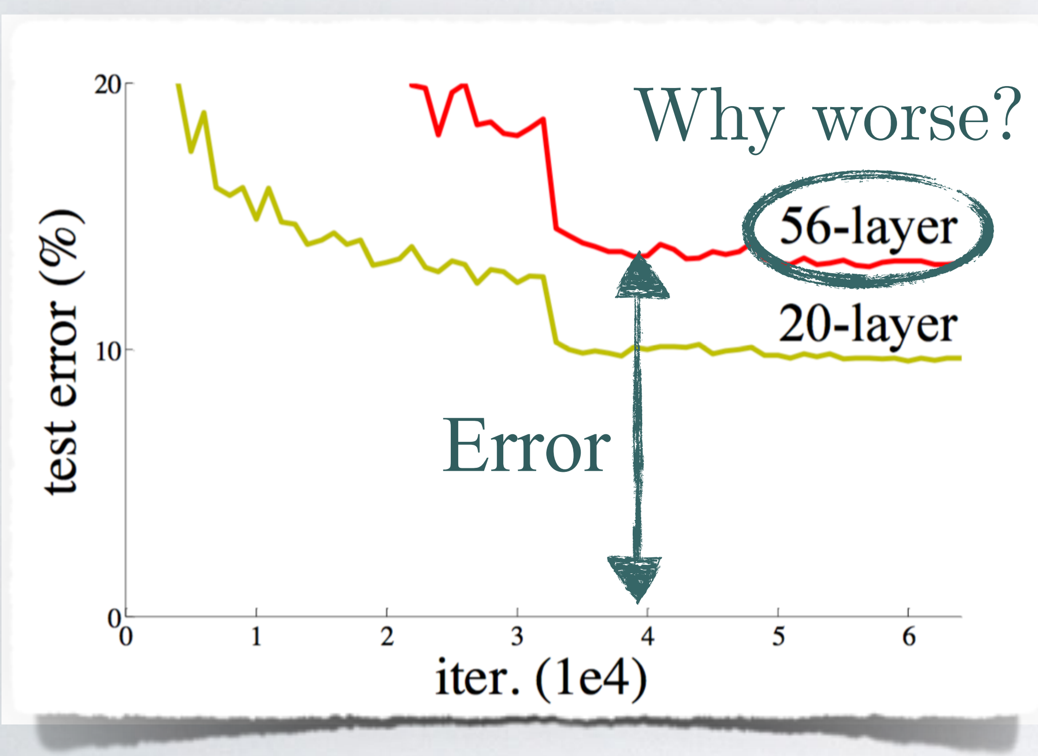


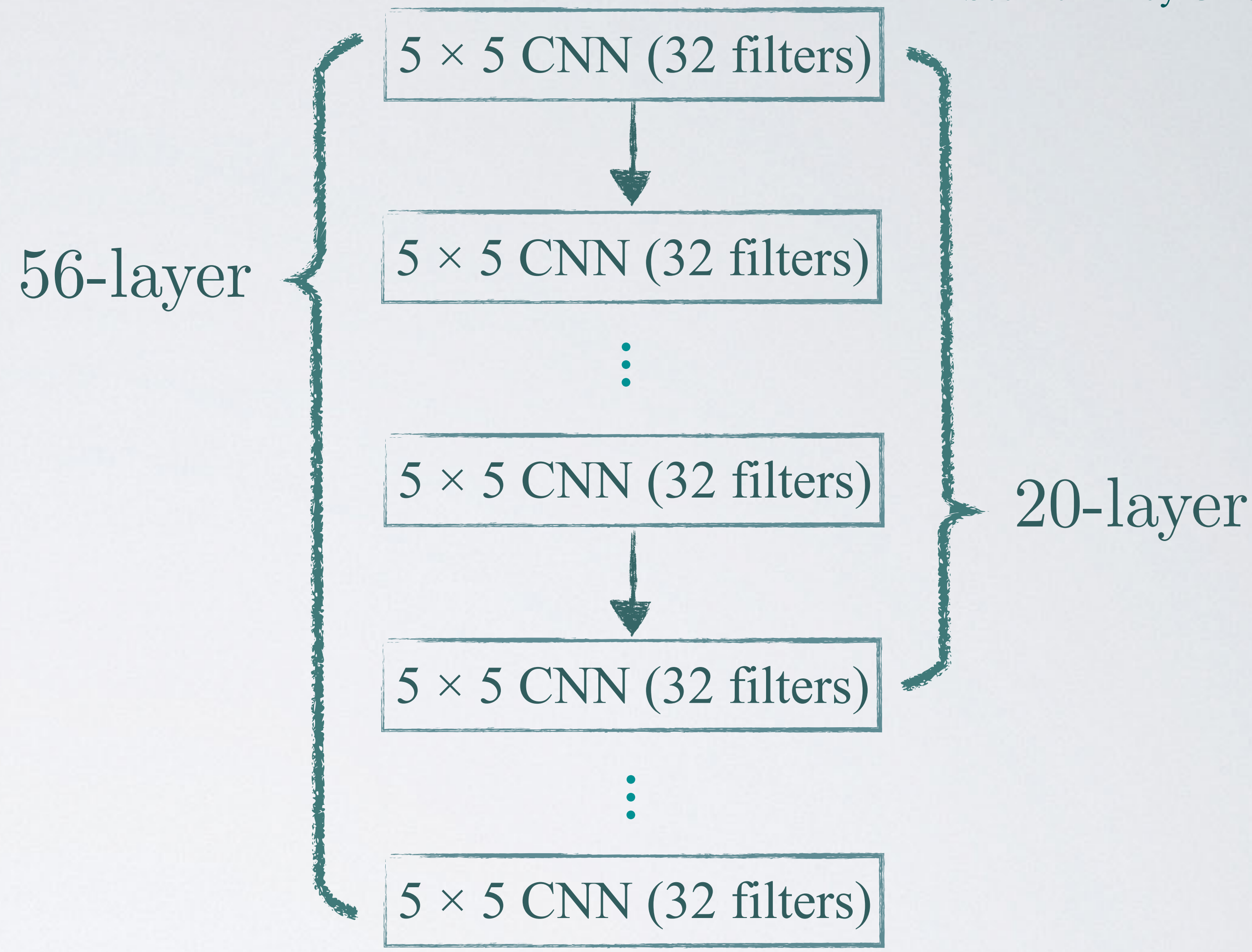
- Deeper neurons are known to be sensitive to shapes.
- Architecturewise, since we have multiple images, we wanted to go deeper. That was our prime obsession.

A Limitation of the Classic CNN

See the talk by Gilles





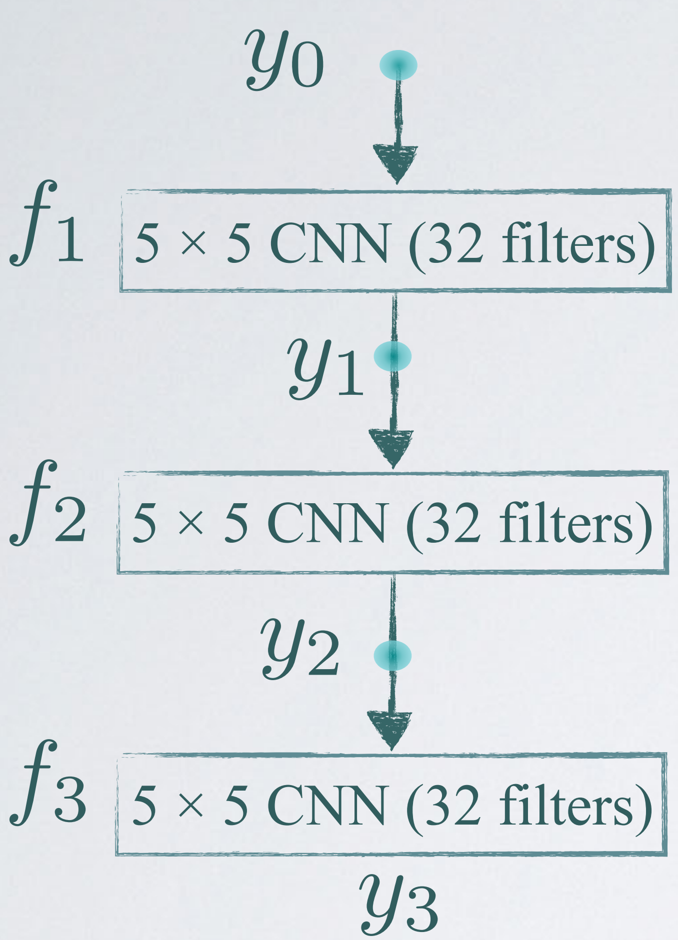


- However, as the network goes deeper, its performance gets saturated or even starts degrading rapidly (vanishing gradient problem).
- Many kinds of alternative CNN were proposed, e.g. Capsule Network, Matrix Capsule Network, … S. Sabour, N. Frosst, G. Hinton [2017]

  S. Diefenbacher, H. Frost, G. Kasieczka, T. Plehn, J. M. Thompson [2019]

A Residual Neural Network (ResNet)

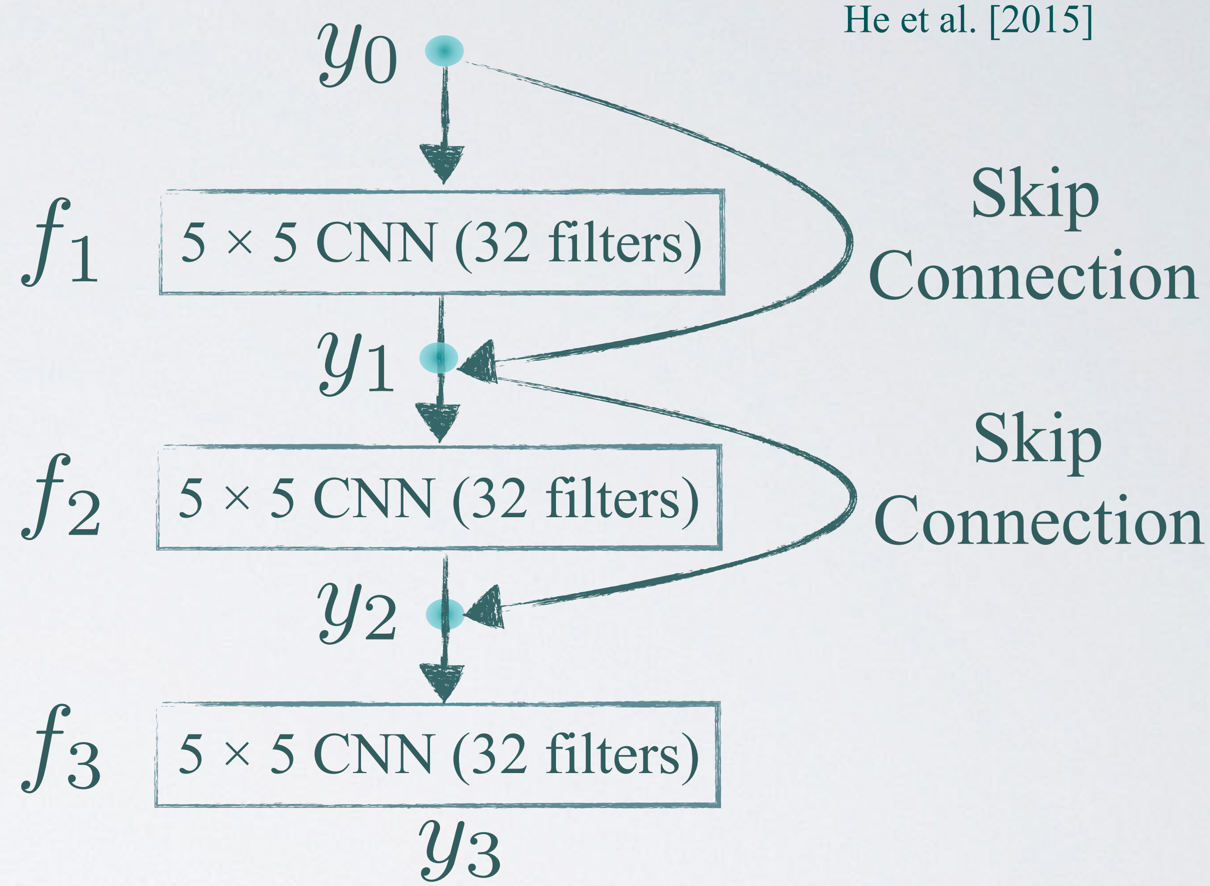
Classic CNN



$$y_3 = f_3(f_2(f_1(y_0)))$$

Residual Neural Network (ResNet)

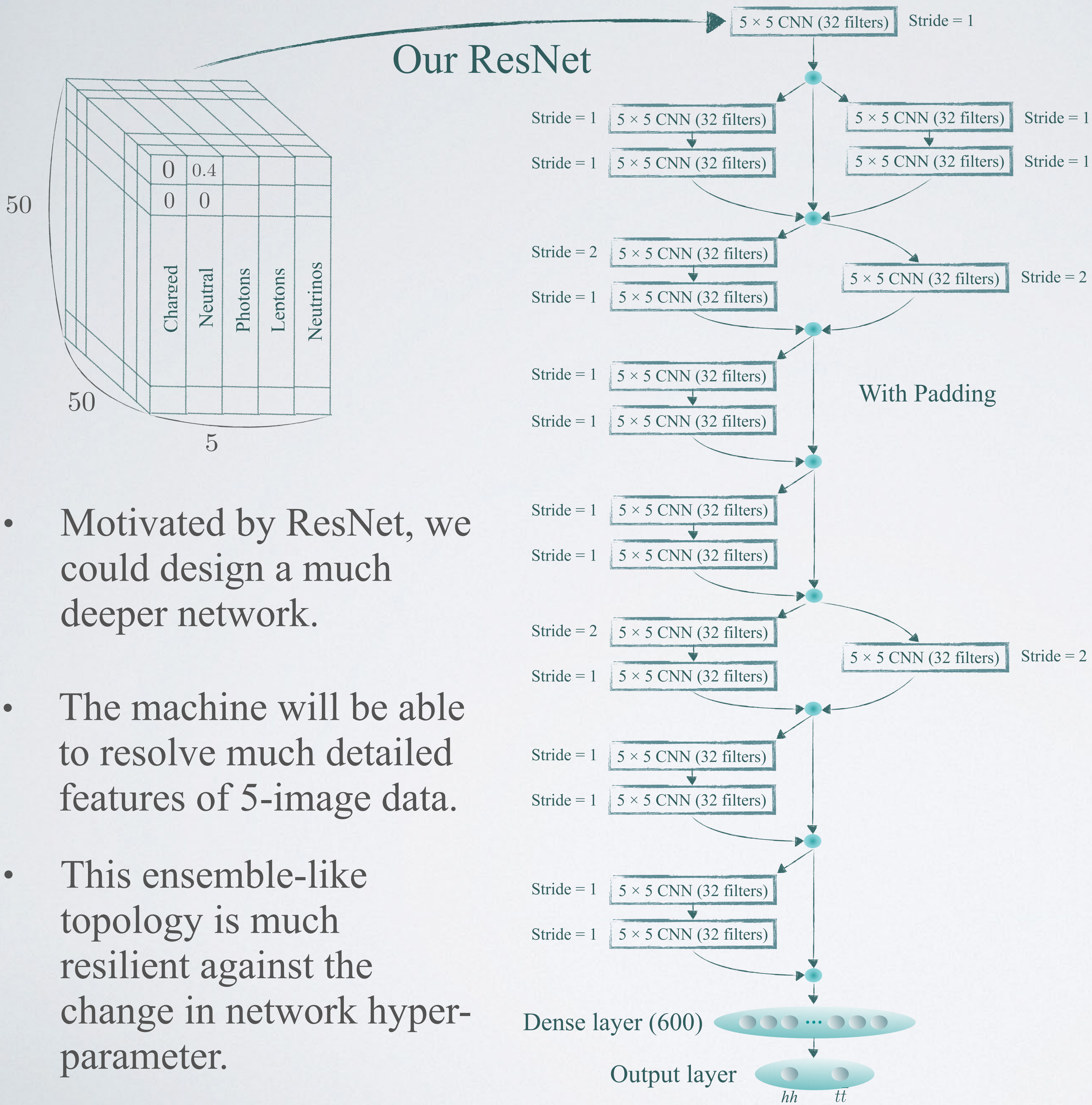
He et al. [2015]



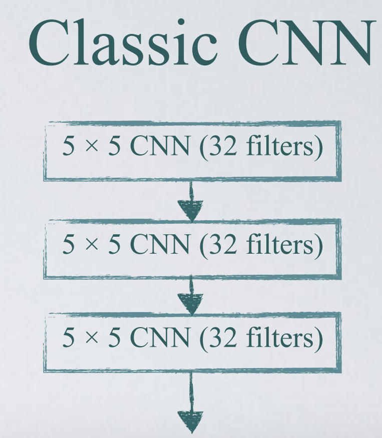
$$y_3 = \underset{\text{original}}{y_0} + \underset{\text{angle1}}{f_1(y_0)} + \underset{\text{angle2}}{f_2(y_0 + f_1(y_0))}$$
$$+ \underset{\text{angle3}}{f_3(y_0 + f_1(y_0) + f_2(y_0 + f_1(y_0)))}$$

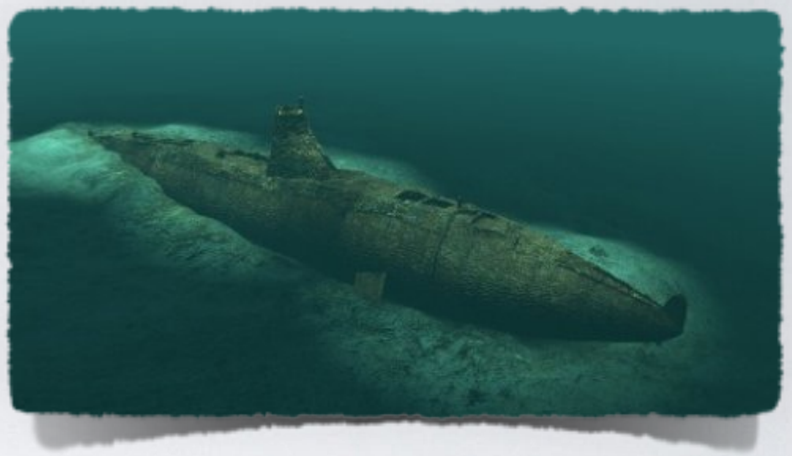
- The ResNet was proposed to go much deeper.
- The core idea of ResNet is introducing skip-connections of nodes.
- ResNet keeps various transformations of images, compared to the classic CNN.

Our ResNet



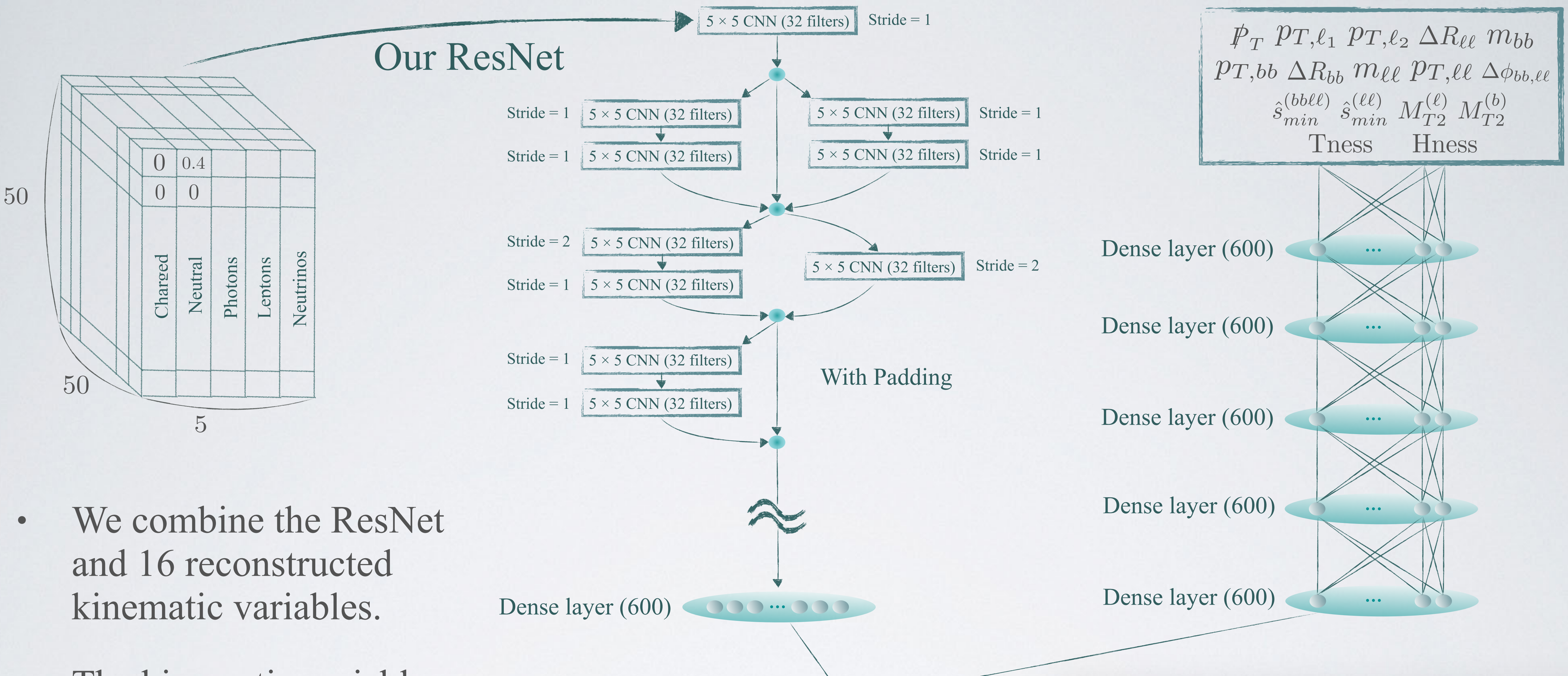
- Motivated by ResNet, we could design a much deeper network.
- The machine will be able to resolve much detailed features of 5-image data.
- This ensemble-like topology is much resilient against the change in network hyper-parameter.





- The classic CNN starts to degrade after 3 layers, with our 5-image data.
- Absolutely not suitable.

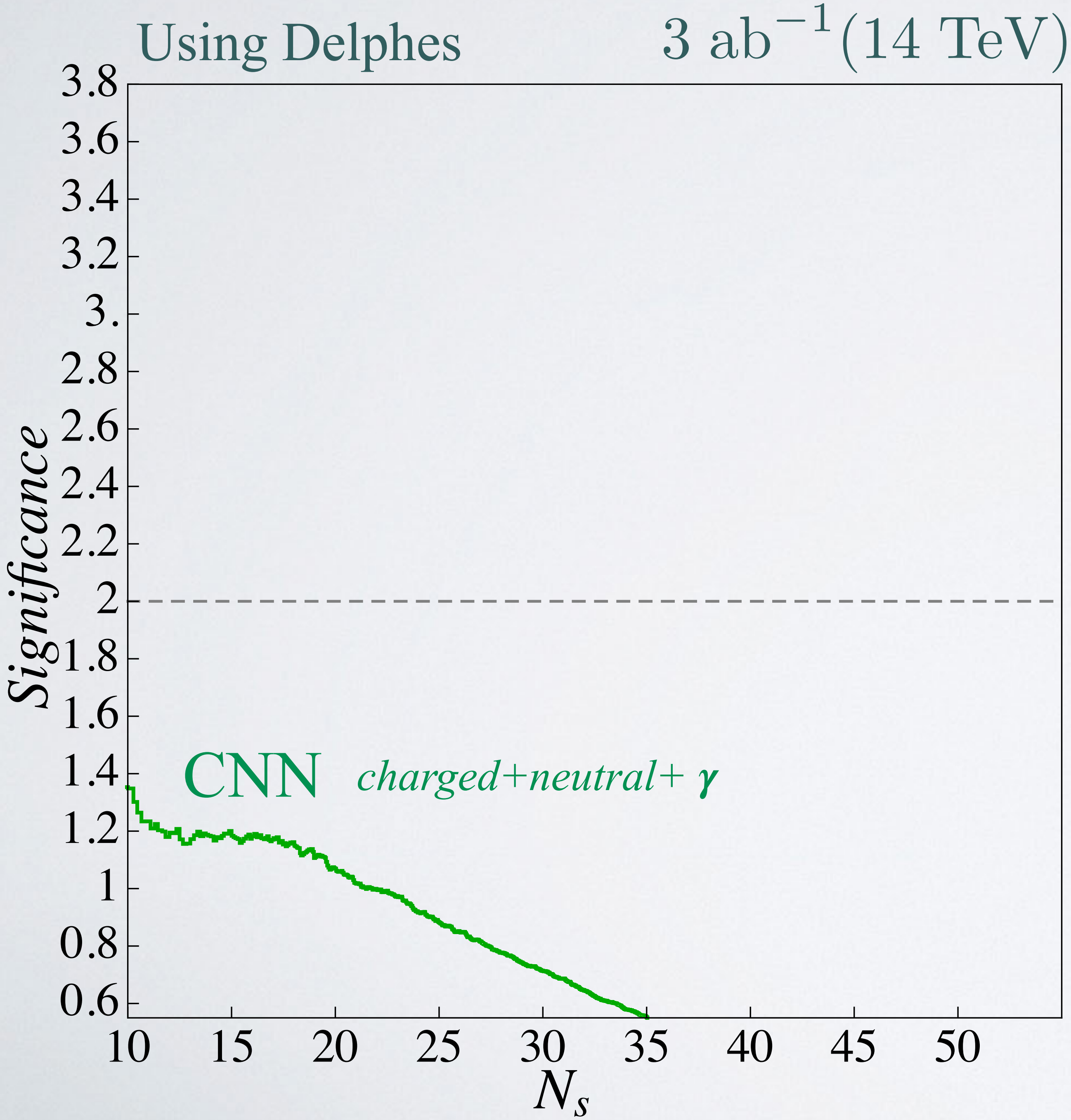
Our ResNet + Kinematic Variables

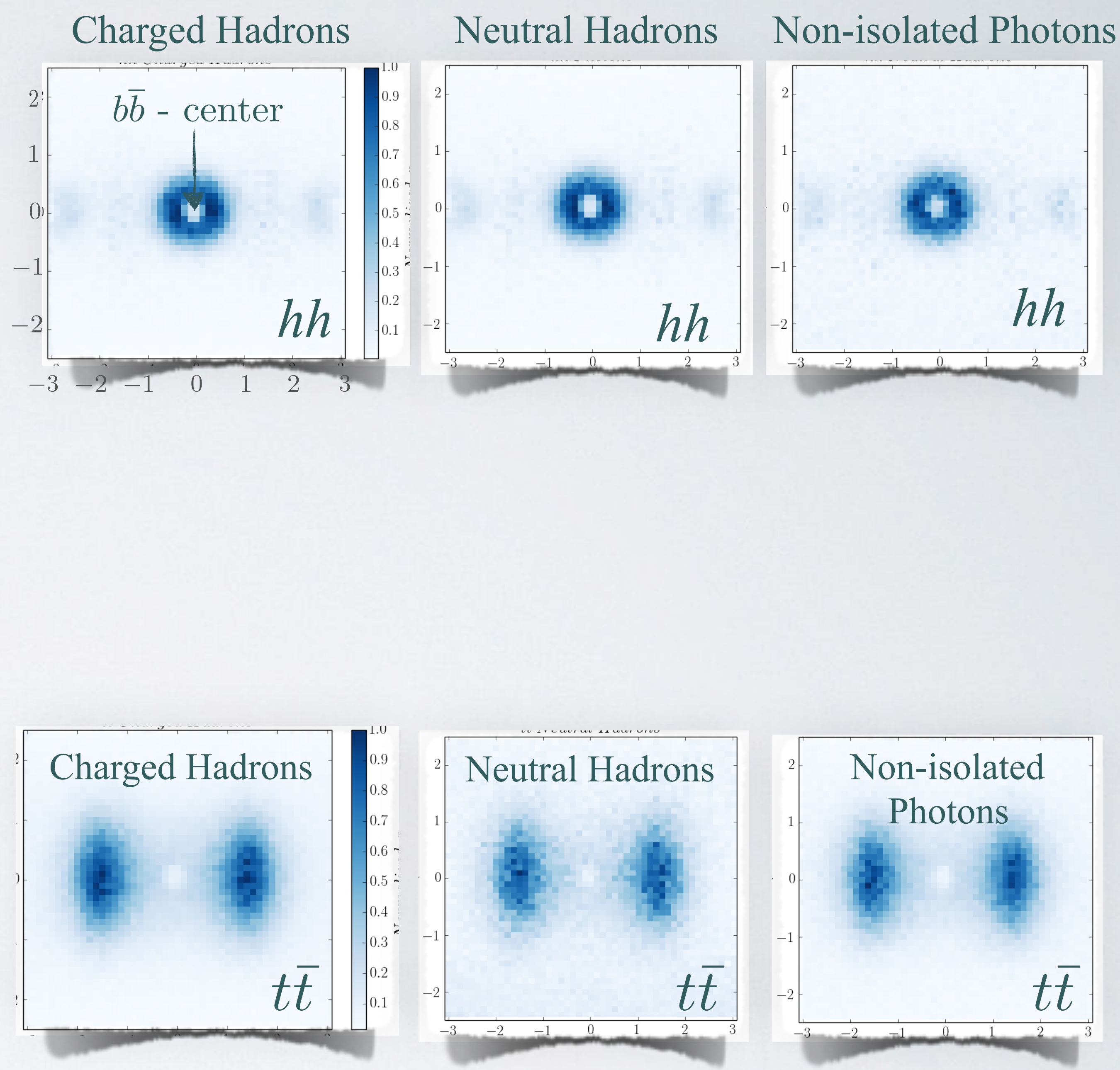


- We combine the ResNet and 16 reconstructed kinematic variables.
- The kinematic variables are processed using fully connected layers.
- Two different architectures are combined in the deep down of the network (complementarity).

Preliminary Results

- We compute the discovery significance of *hh* with 3 ab^{-1}
- First we utilize the classic version of CNN with only three images.



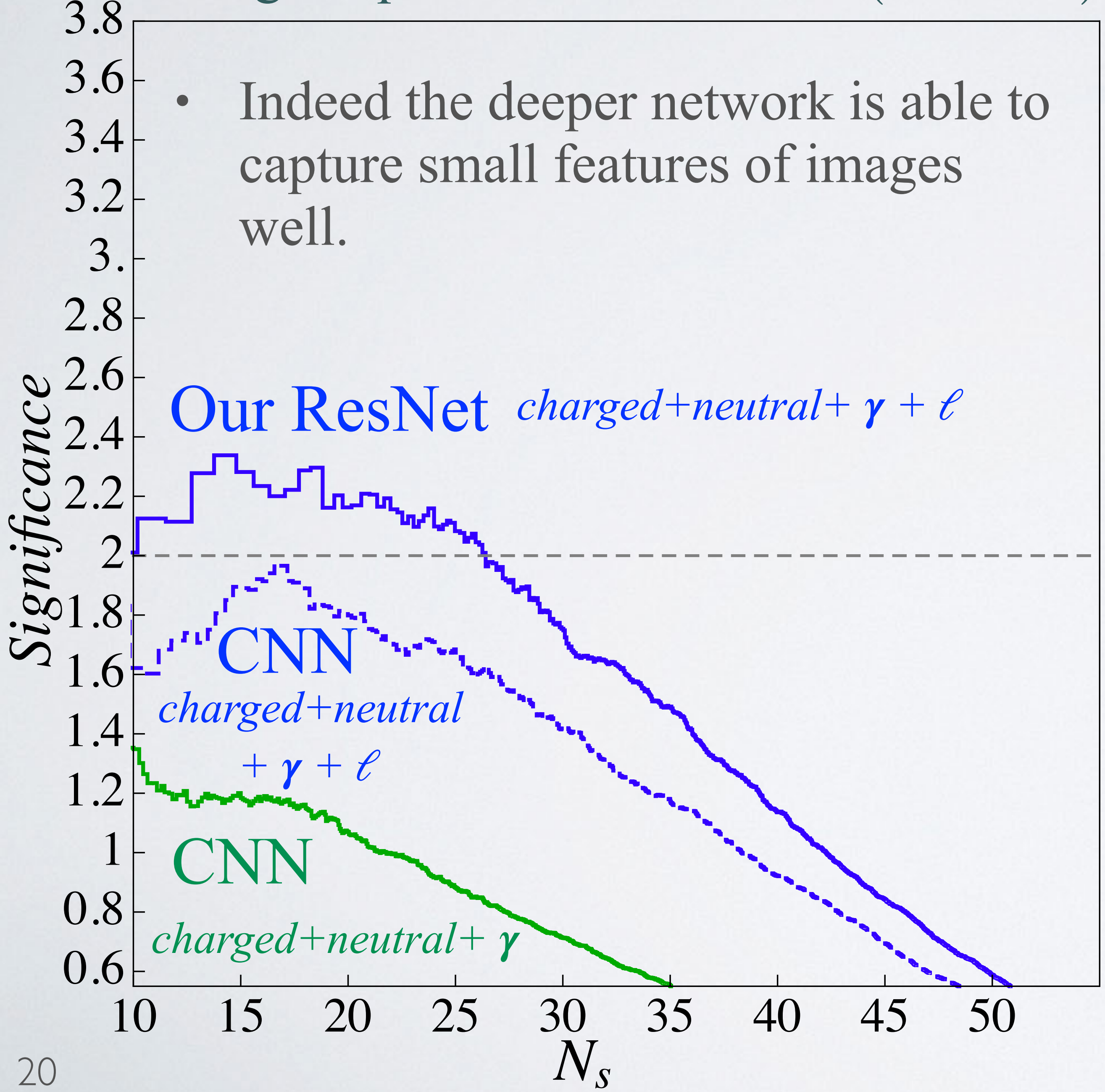


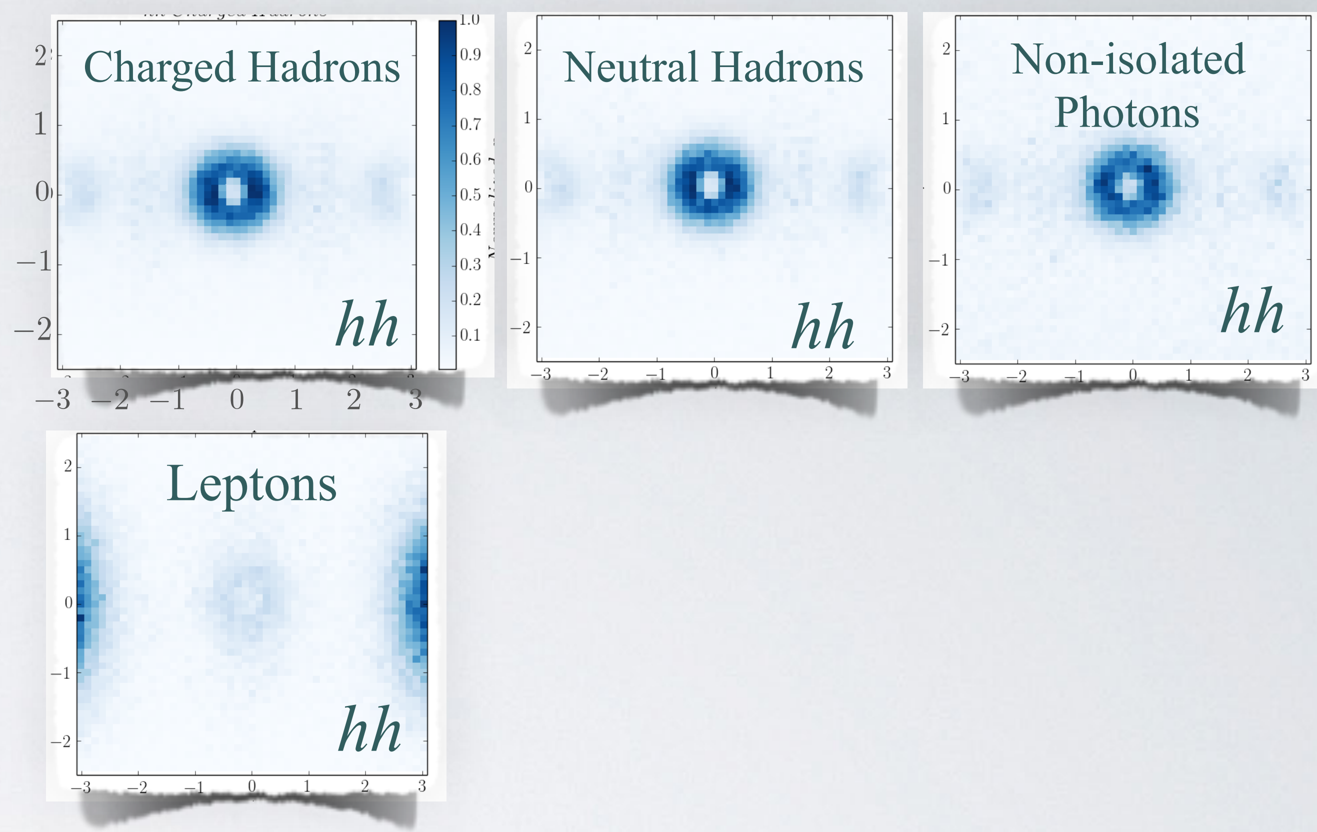
Preliminary Results

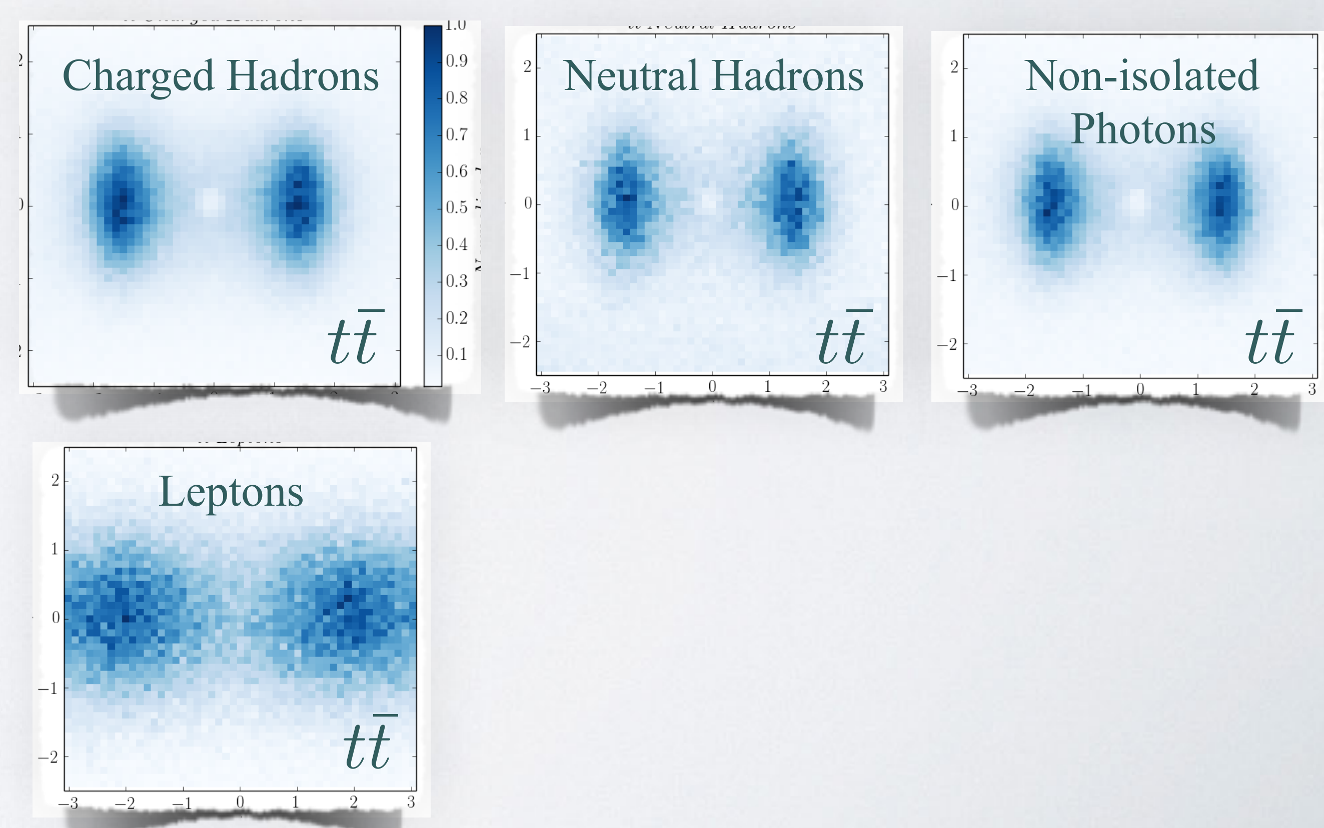
- Lepton images provide an orthogonal information.
- Adding lepton images with our ResNet drastically improves the result.

Using Delphes $3\ \mathrm{ab}^{-1}(14\ \mathrm{TeV})$

- Indeed the deeper network is able to capture small features of images well.

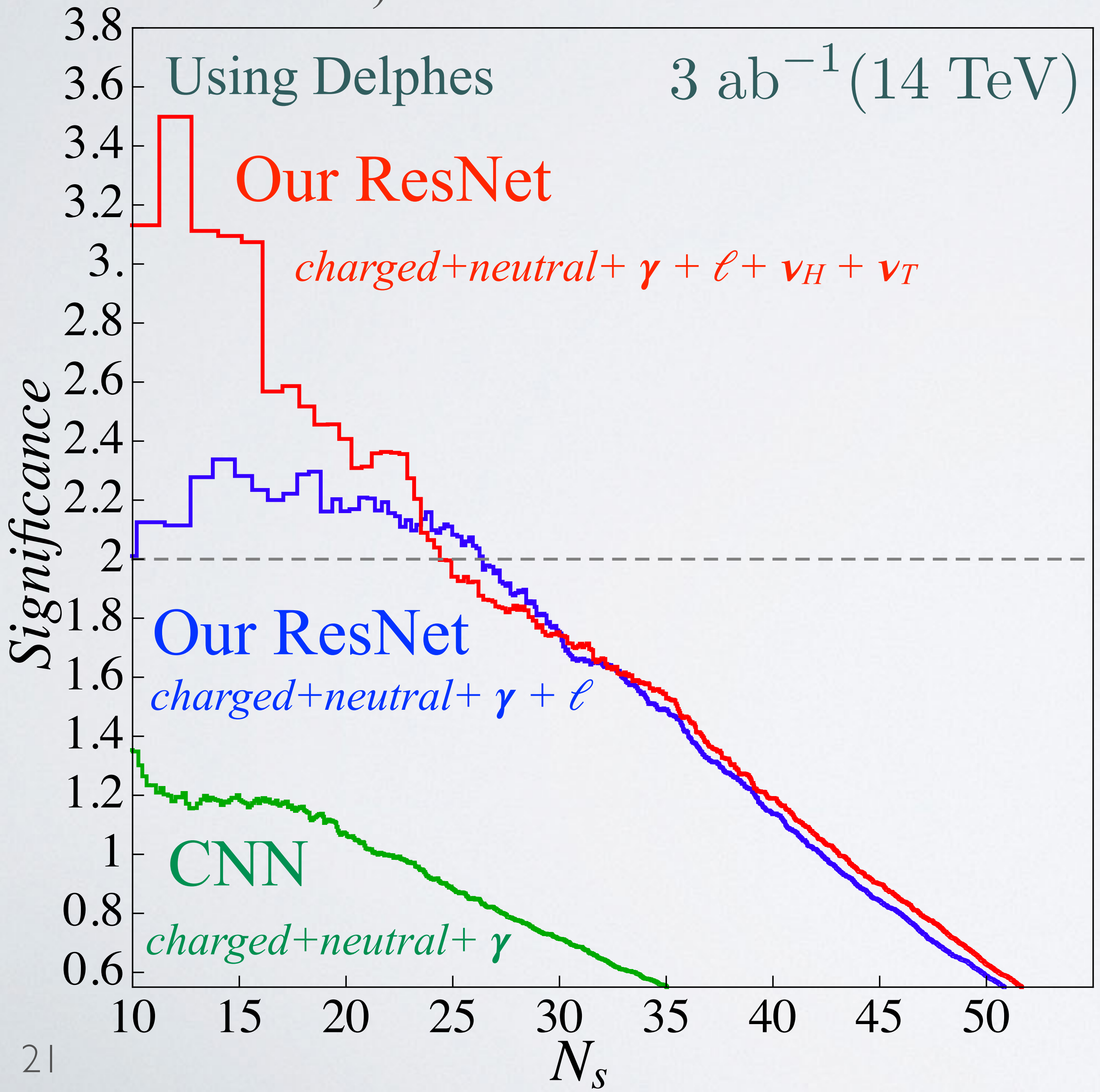






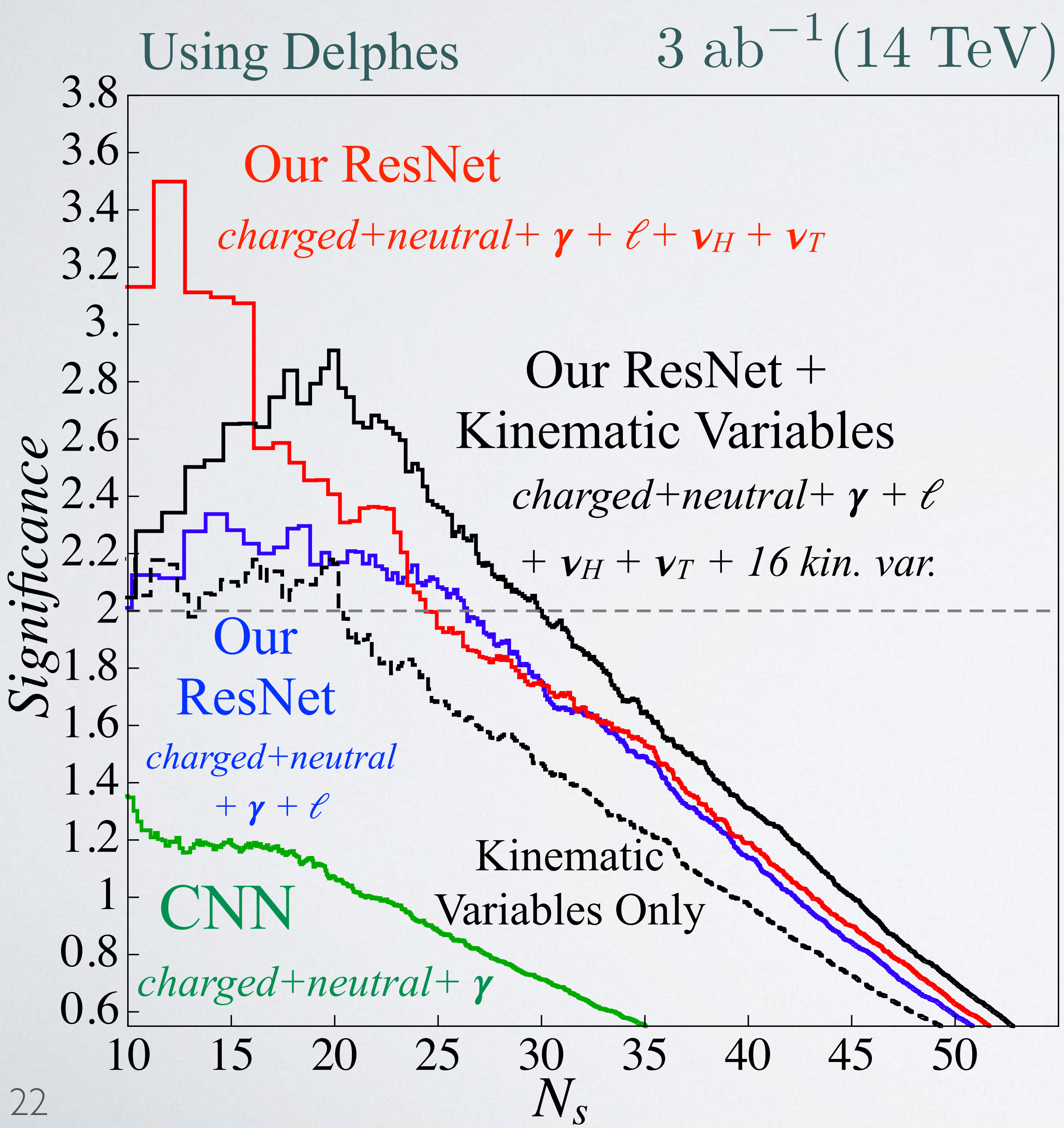
Preliminary Results

- Reconstructed neutrino images have large correlations with lepton images.
- Overall performance doesn't change much.
- The ResNet starts to see the difference in the low efficiency region (vulnerable to statistics)



Preliminary Results

- Combining kinematic variables improves the result in the wide range of signal region.
- Two information is **complementary**



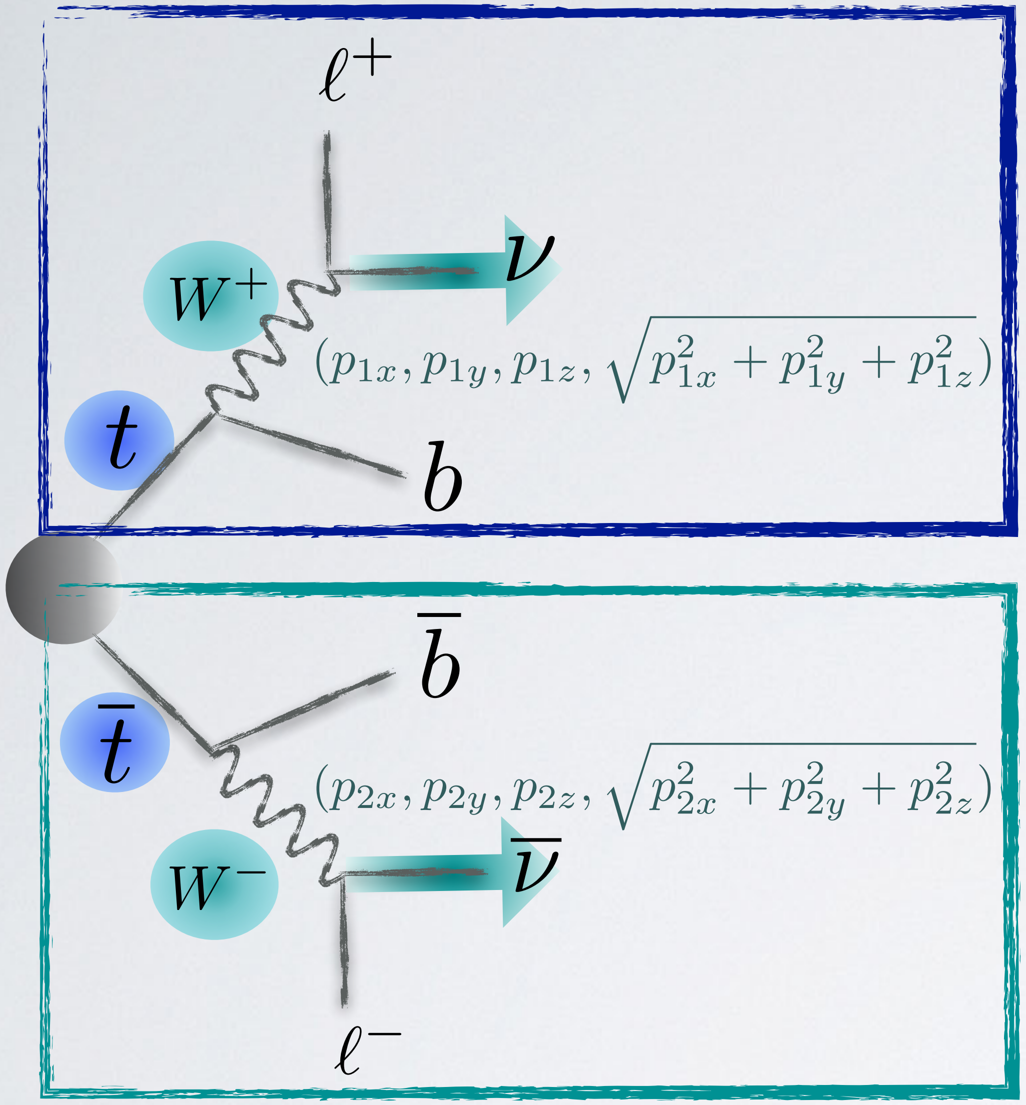
“Hence the purely image-based neural network has a potential to improve the previous results. It’s not a factor of 2 improvement, but it may be just what we need to discover the *hh* when combining all channels.”

⬇ back-up ⬇

Topness (T)

M. L. Graesser and J. Shelton [2012]

J. H. Kim, K. C. Kong, K. T. Matchev, M. Park [2018]



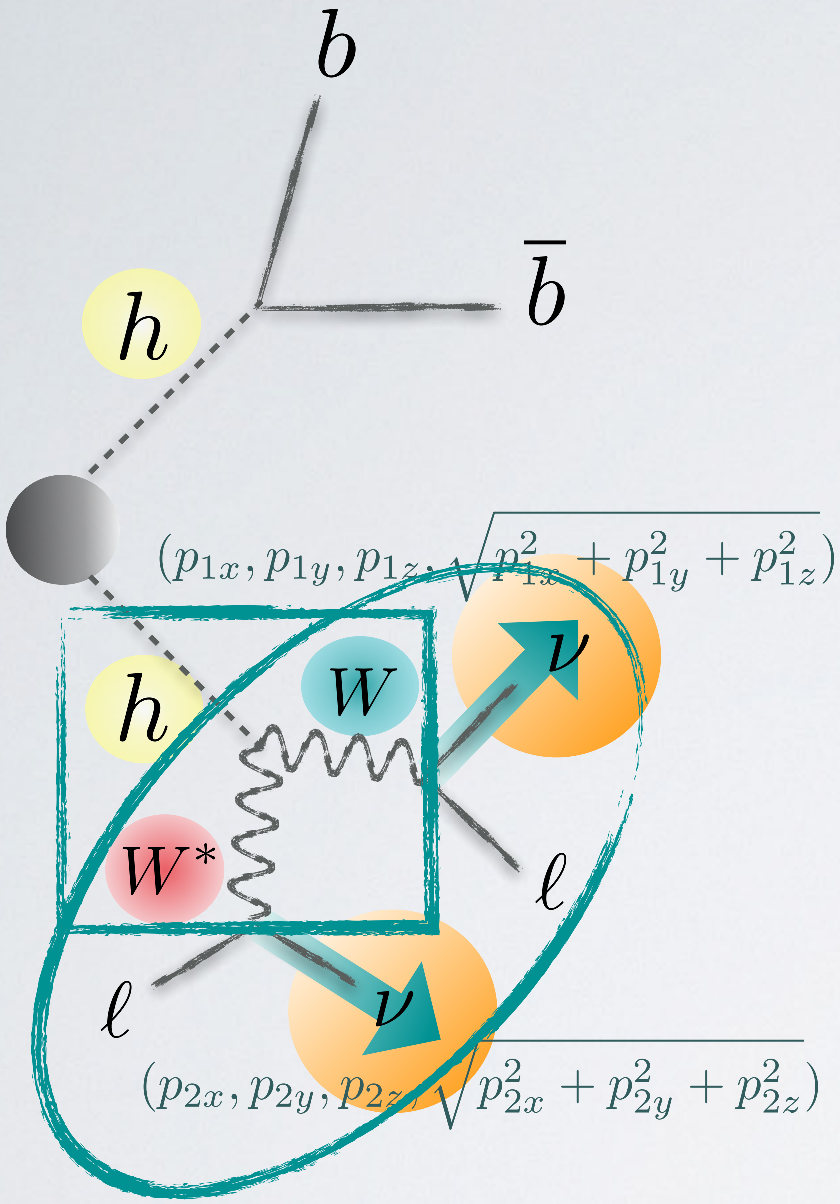
$$\chi_{ij}^2 \equiv \min_{\vec{p}\!\!/_T = \vec{p}_{\nu T} + \vec{p}_{\bar{\nu} T}} \left[\frac{\left(m_{b_i \ell^+ \nu}^2 - m_t^2\right)^2}{\sigma_t^4} + \frac{\left(m_{\ell^+ \nu}^2 - m_W^2\right)^2}{\sigma_W^4} + \frac{\left(m_{b_j \ell^- \bar{\nu}}^2 - m_t^2\right)^2}{\sigma_t^4} + \frac{\left(m_{\ell^- \bar{\nu}}^2 - m_W^2\right)^2}{\sigma_W^4} \right]$$

$$\boxed{T \equiv min\left(\chi_{12}^2\, ,\, \chi_{21}^2\right)}$$

two possible ways of paring b and ℓ

- Topness provides a degree of consistency to dileptonic $t\bar{t}$ production.
- It scans over 6 unknowns of neutrino momenta with four on-shell masses and missing E_T constraints.
- And find a minimum of the likelihood function.

Higgsness (*H*)



J. H. Kim, K. C. Kong, K. T. Matchev, M. Park [2018]

$$H \equiv \min_{\vec{\not p}_T = \vec{p}_{\nu T} + \vec{p}_{\bar\nu T}} \left[\boxed{\frac{\left(m^2_{\ell^+\ell^-\nu\bar\nu} - m_h^2\right)^2}{\sigma^4_{h_\ell}}} + \frac{\left(m^2_{\nu\bar\nu} - m^2_{\nu\bar\nu,peak}\right)^2}{\sigma^4_\nu} \right.$$

$$+ min \left(\boxed{\frac{\left(m^2_{\ell^+\nu} - m_W^2\right)^2}{\sigma^4_W} + \frac{\left(m^2_{\ell^-\bar\nu} - m^2_{W^*,peak}\right)^2}{\sigma^4_{W_*}}}, \right.$$

$$\left. \left. \boxed{\frac{\left(m^2_{\ell^-\bar\nu} - m_W^2\right)^2}{\sigma^4_W} + \frac{\left(m^2_{\ell^+\nu} - m^2_{W^*,peak}\right)^2}{\sigma^4_{W_*}}} \right) \right],$$

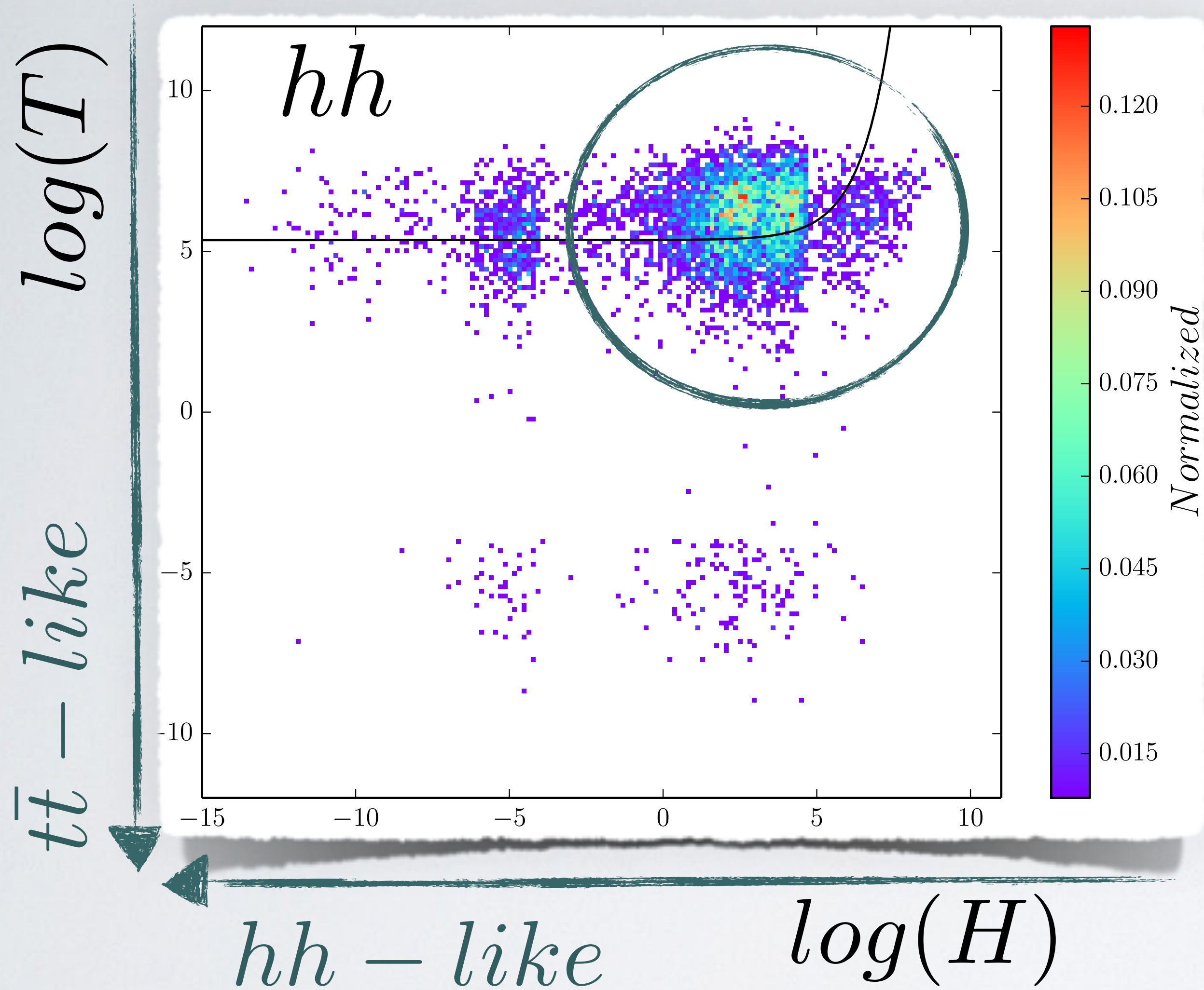
two possible ways of paring ν and ℓ

$\sim m_h - m_W$
off-shell

- Higgsness provides a degree of consistency to dileptonic $h \to WW^*$ system.
- The off-shell *W* also has an end-point near m_h - m_W.
- Its distribution is wide, but there is a peak, which can constrain *hh* system further.

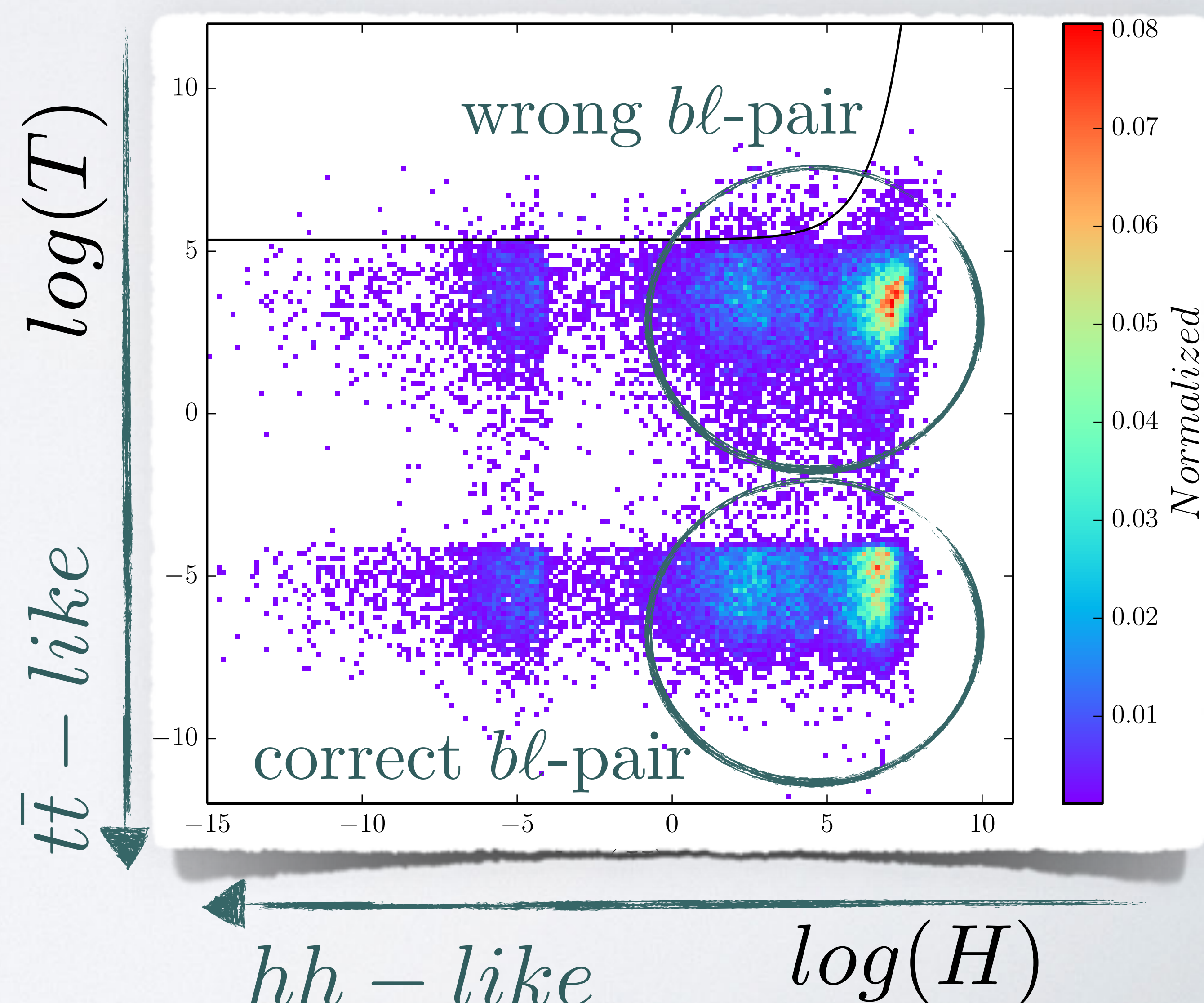
Distributions of (*log H* , *log T*) after baseline selection cuts

J. H. Kim, K. C. Kong, K. T. Matchev, M. Park [2018]



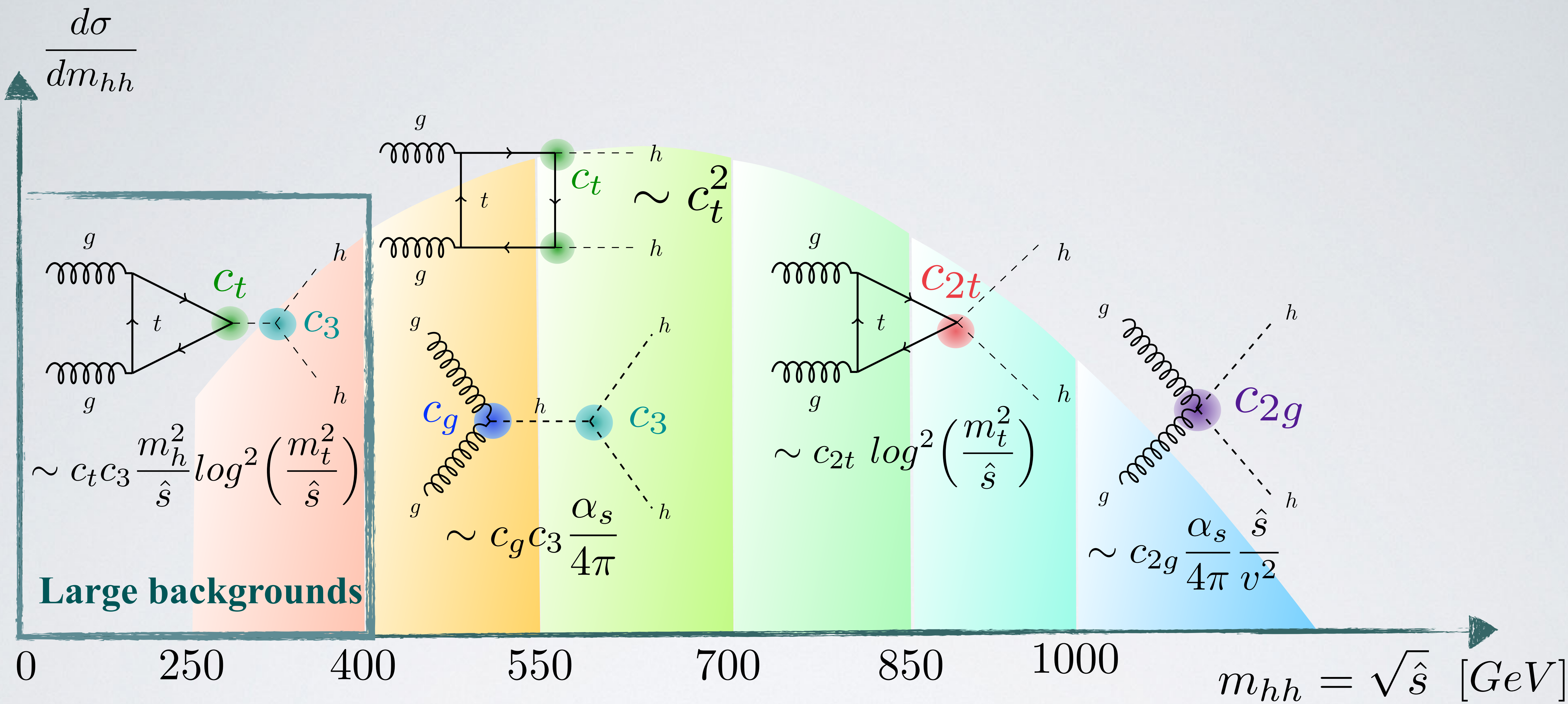
- A clear separation between *hh* and backgrounds ($t\bar{t}$ is dominant)

All backgrounds



- Since there is a two-fold ambiguity in $b\ell$-paring, Topness displays the island-nature.

Energy-dependence & Shape analysis



- Interesting thing about these couplings in *hh* is that they all display different Energy-dependences.
- The c_3 is sensitive at lower-energy bins. That's why it's hard to measure!
- At higher-energy bins, c_{2g} and c_{2t} become more sensitive.

The Effective Lagrangian for *hh*

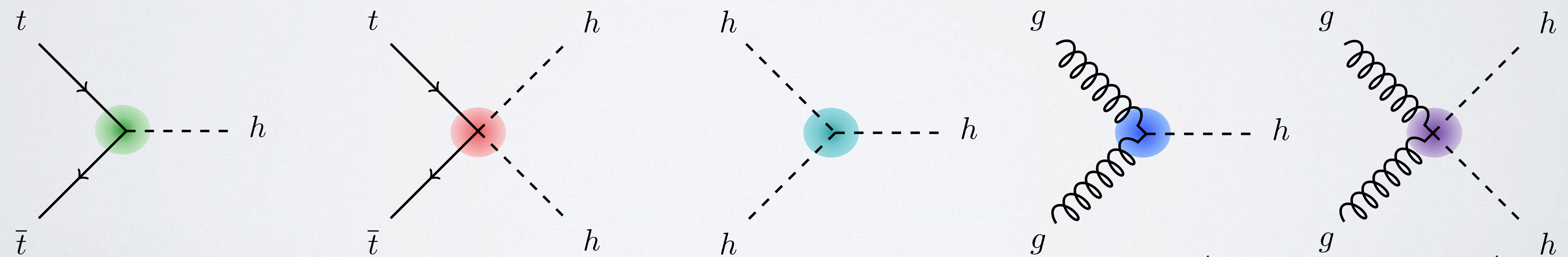
$$\mathcal{L} = \mathcal{L}_{\rm SM} + \boxed{\Delta\mathcal{L}_6} + \Delta\mathcal{L}_8 + \cdots$$

- The resulting vertices are highly correlated with each other…

$$\frac{\bar{c}_H}{2v^2}(\partial^\mu |H|^2)^2 - \frac{\bar{c}_6}{v^2}\left(\frac{m_h^2}{2v^2}\right)|H|^6 + \frac{\bar{c}_u}{v^2}\mathrm{y}_t(|H|^2\overline{Q}_L H^c t_R + \mathrm{h.c.}) + \bar{c}_g\left(\frac{g_s^2}{m_w^2}\right)|H|^2 G^a_{\mu\nu}G^{a\,\mu\nu}$$

After EWSB…

$$\mathcal{L} = -m_t t\bar{t}\left(c_t\frac{h}{v} + c_{2t}\frac{h^2}{v^2}\right) - \frac{c_3}{6}\left(\frac{3m_h^2}{v}\right)h^3 + \frac{g_s^2}{4\pi^2}\left(c_g\frac{h}{v} + c_{2g}\frac{h^2}{2v^2}\right)G^a_{\mu\nu}G^{a\mu\nu} + \cdots$$



$$c_t \simeq 1 - \frac{\bar{c}_H}{2} - \bar{c}_u \qquad c_{2t} \simeq -\frac{1}{2}(\bar{c}_H + 3\bar{c}_u) \qquad c_3 \simeq 1 - \frac{3}{2}\bar{c}_H + \bar{c}_6 \qquad c_g \simeq \bar{c}_g\left(\frac{4\pi}{\alpha_2}\right) \qquad c_{2g} \simeq \bar{c}_g\left(\frac{4\pi}{\alpha_2}\right)$$

(where $\alpha_2 \equiv g^2/4\pi$)