Human Age Estimation
Based on Trainable Gabor wavelet

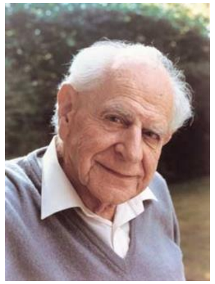
Hyuk Jin Kwon, Jae Woong Soh, Hyung Il Koo, **Nam Ik Cho**

SEOUL
NATIONAL
UNIVERSITY

Introduction

What is Human Age Estimation?

- **As a Computer Vision Task**
  - Determine appearance age of a human from his near frontal photographs

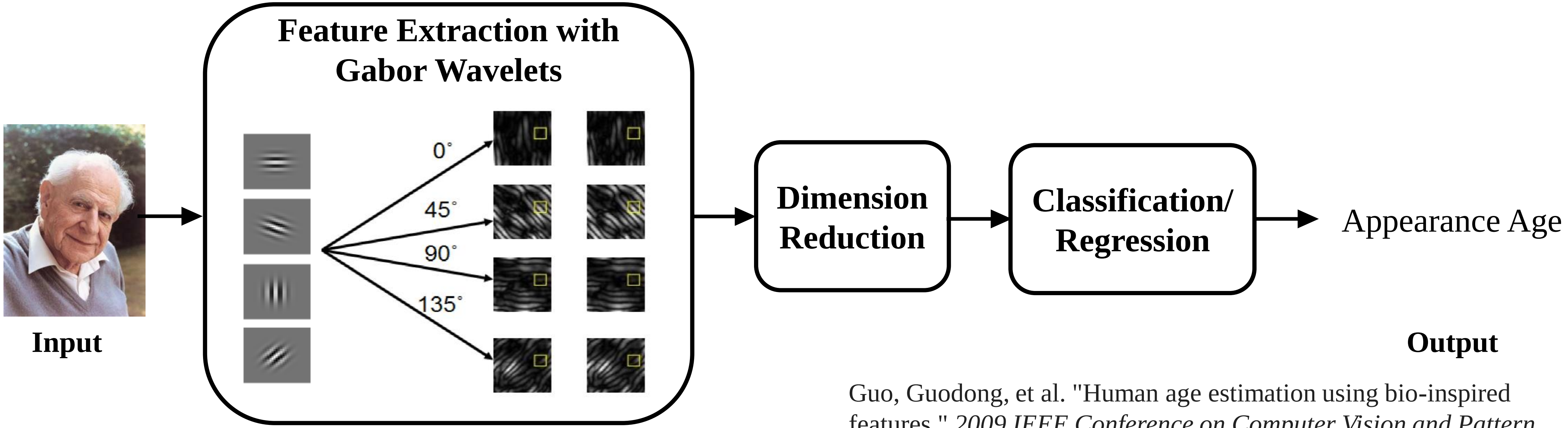


Input → Classification or Regression → 90 years old

Output

Existing Methods

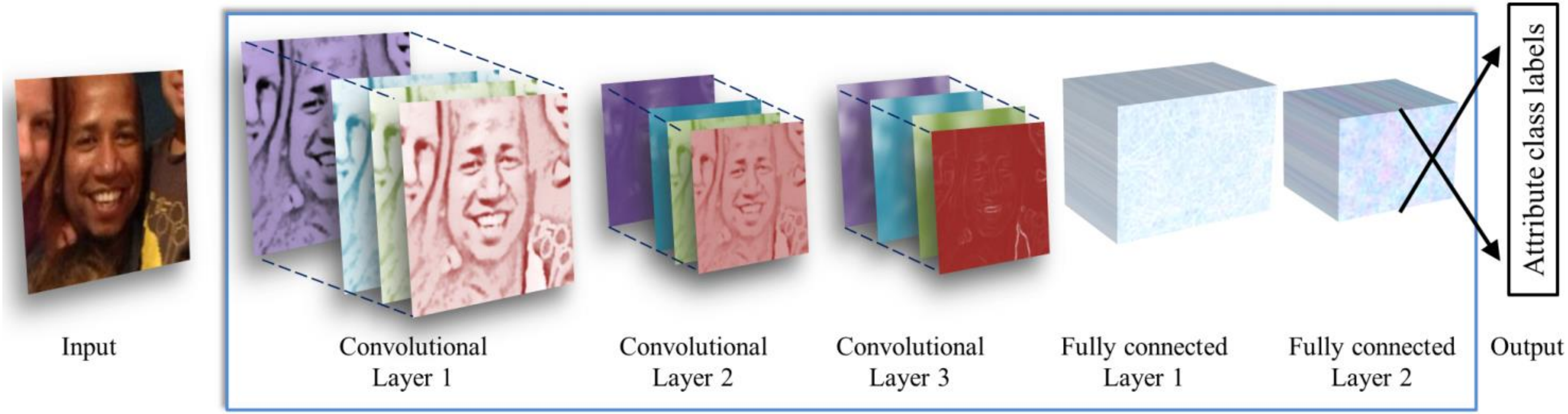
- **Feature Point Based Classifiers (Before 2012)**
  - Gabor Wavelet Based Methods



Guo, Guodong, et al. "Human age estimation using bio-inspired features." *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009.

Existing Methods

- **Convolutional Neural Network (After 2012)**



Levi, Gil, and Tal Hassner. "Age and gender classification using convolutional neural networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015.

Existing Methods

	Feature Based Classifiers	Convolutional Neural Network
Pros	• Smaller number of parameters	• High Performance • Fully trainable
Cons	• Low performance • Manual tuning of hyper-parameters	• Larger number of parameters

Existing Methods

	Feature Based Classifiers	Convolutional Neural Network
Pros	• Smaller number of parameters	• High Performance • Fully trainable
Cons	• Low performance • Manual tuning of hyper-parameters	• Larger number of parameters

Existing Methods

	Feature Based Classifiers	Convolutional Neural Network
Pros	• Smaller number of parameters	• High Performance • Fully trainable
Cons	• Low performance • Manual tuning of hyper-parameters	• Larger number of parameters

What We Want

	Feature Based Classifiers	Convolutional Neural Network
Pros	• Smaller number of parameters	• High Performance • Fully trainable
~~**Cons**~~	~~• Low performance~~ ~~• Manual tuning of hyper-parameters~~	~~• Larger number of parameters~~

Our Approach

Why we used Gabor wavelets?

- **Requires Small Number of Parameters:**

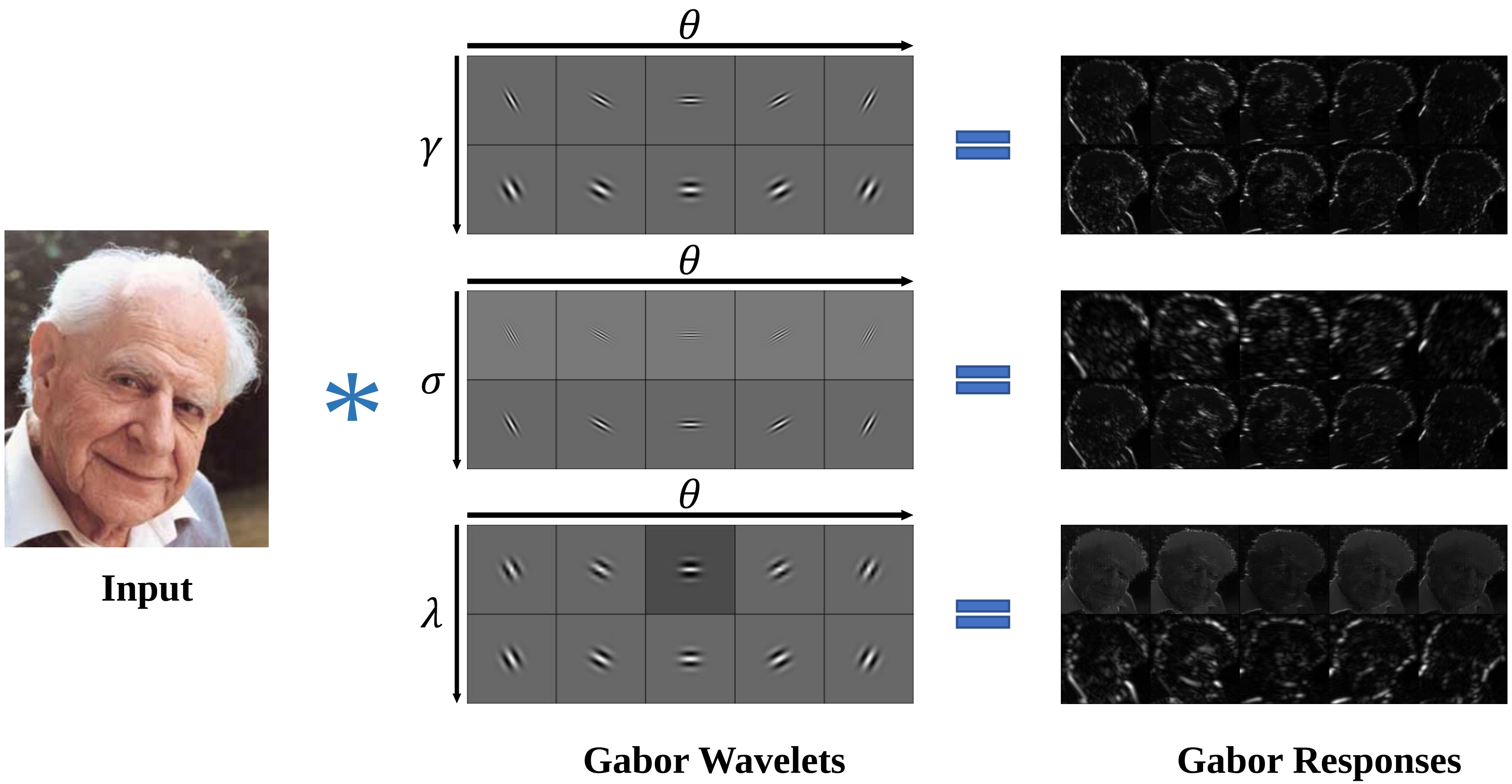
  2D Gabor wavelets can be obtained with only 4 parameters $(\gamma, \sigma, \lambda, \theta)$ and sampling grid.

$$G(x,y) = \exp\left(-\frac{X^2 + \boxed{\gamma}^2 Y^2}{2\boxed{\sigma}^2}\right) \times \cos\left(\frac{2\pi}{\boxed{\lambda}} X\right)$$

$$X = x\cos\boxed{\theta} + y\sin\theta, Y = -x\sin\theta + y\cos\theta$$ **+ Sampling Grid**

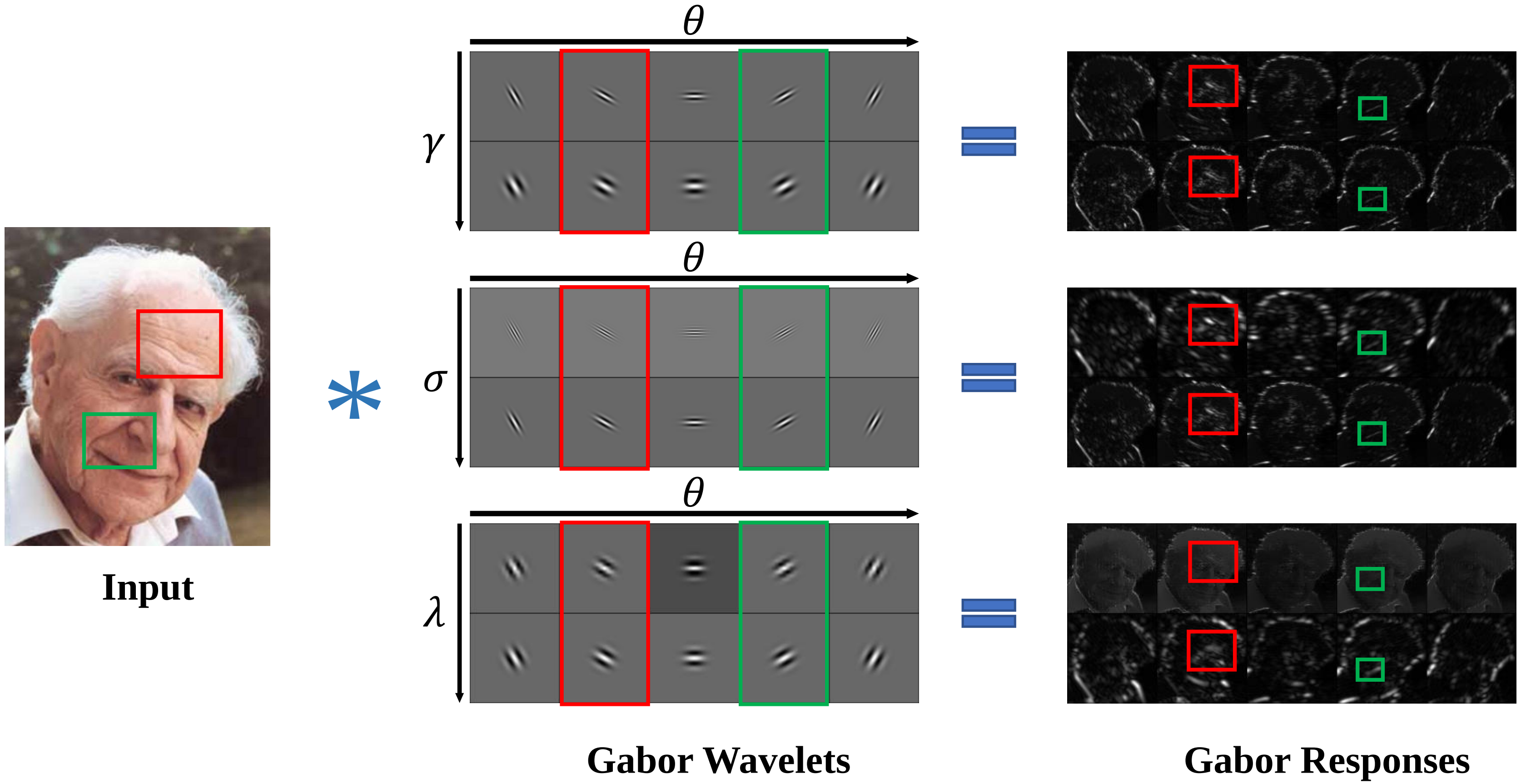
Why we used Gabor wavelets?

- **Feature Extractor that Effective in Textural information:**



Why we used Gabor wavelets?

- **Feature Extractor that Effective in Textural information:**



What We Want

	Feature Based Classifiers	Convolutional Neural Network
Pros	• Smaller number of parameters	• High Performance • Fully trainable
~~**Cons**~~	~~• Low performance~~ ~~• Manual tuning of parameters~~	~~• Larger number of parameters~~

What We Want

	Feature Based Classifiers	Convolutional Neural Network
Pros	• Smaller number of parameters	• High Performance • Fully trainable ?
~~**Cons**~~	~~• Low performance~~ ~~• Manual tuning of parameters~~	~~• Larger number of parameters~~

Challenges

How can we incorporate Gabor wavelets into CNN in the fully trainable manner?

Estimation of Parameters of Gabor wavelets

- We can manipulate the characteristic of discrete 2D Gabor wavelet by its parameters (γ, σ, λ, θ) and the size of sampling grid

- It means that we can train Gabor wavelets by estimating its parameters and the sampling grid size.

$$G(x,y) = \exp\left(-\frac{X^2 + \gamma^2 Y^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda}X\right)$$

Equation of 2D Gabor wavelets

Modeling of Sampling Grid with ζ

- Intrinsic parameter γ, σ, λ has mathematical relationship to Gabor wavelet
- We need to digitize the Gabor wavelet for actual implementation
- However, the sampling grid size has no connection with the equation of Gabor wavelet
- So **we introduce a new parameter ζ** to make a connection between sampling grid size and Gabor wavelet

Modeling of Sampling Grid with ζ

$$G[m,n] = g(u,v) = \left(\frac{m}{\lfloor\zeta\rfloor}\times\zeta, \frac{n}{\lfloor\zeta\rfloor}\times\zeta\right)$$

m and n are in $-\lfloor\zeta\rfloor, -\lfloor\zeta\rfloor+1, \cdots, \lfloor\zeta\rfloor$

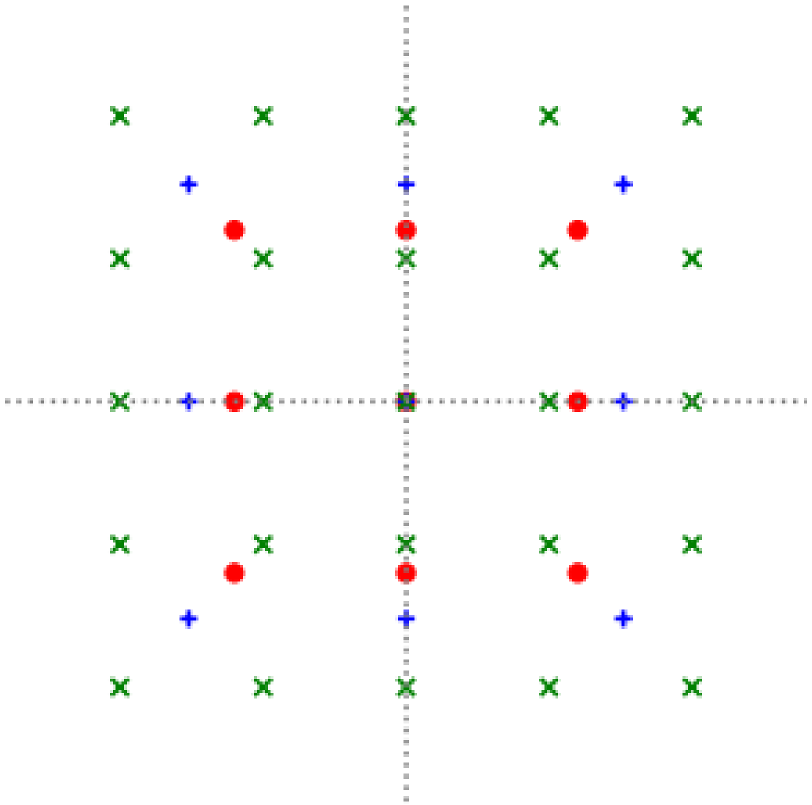
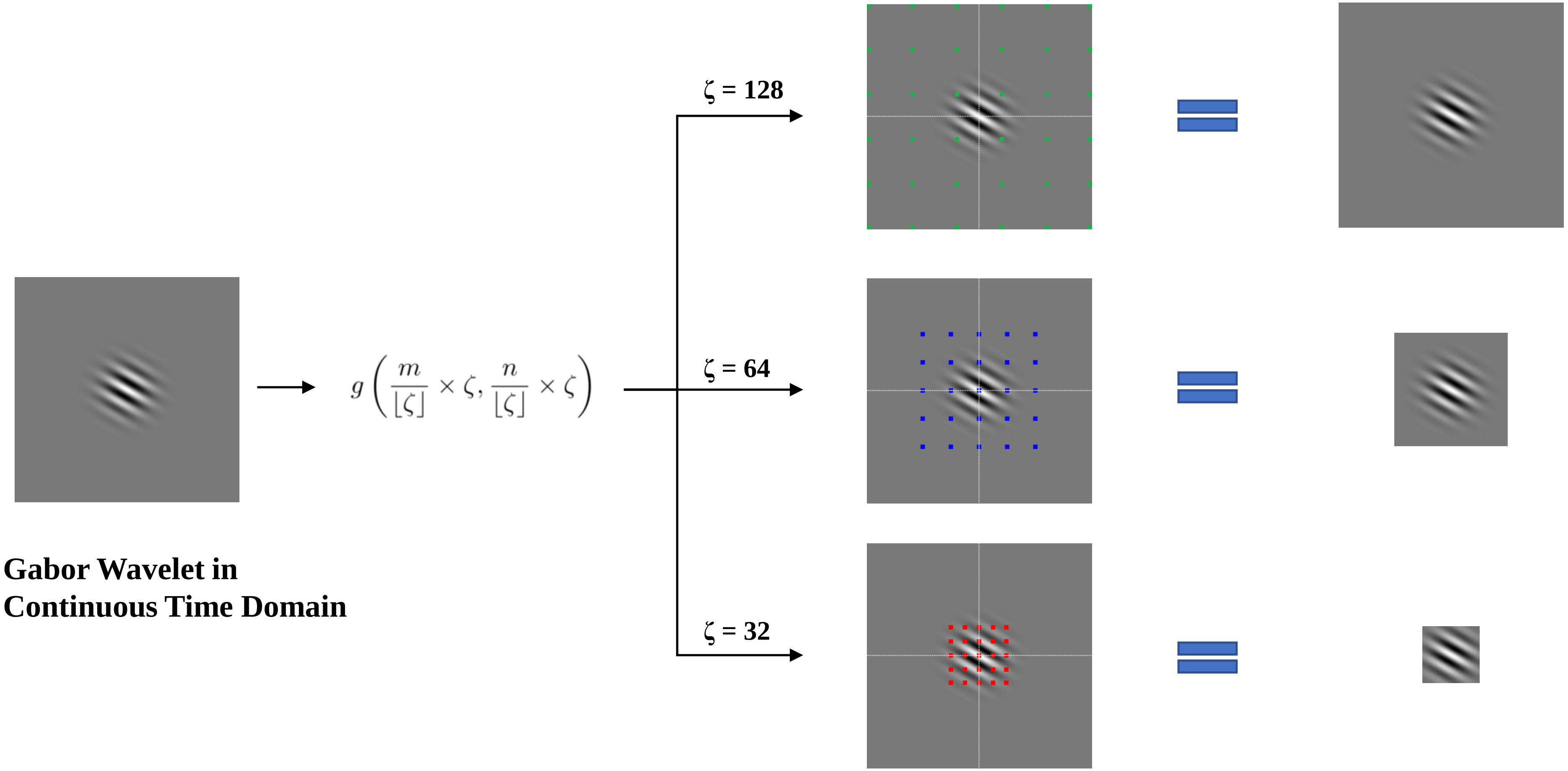


Fig. 2. Sampling grids for three ζ values. Red, blue and green dots are generated by (2) with $\zeta = 1.5$, 1.9 and 2.5 respectively.

Modeling of Sampling Grid with ζ



Gabor Wavelet in
Continuous Time Domain

Sampling of Gabor wavelet
according to ζ

Gabor Wavelet in
Discrete Time Domain

Estimation of Parameters of Gabor wavelets

- CNN with 3x3 convolution was used for estimating parameters γ, σ, λ, and ζ because

  1. CNN with 3x3 convolution is the smallest one among practical CNN

  2. It is fully trainable

Estimation of Parameter θ

- The one last parameter to be determined is the orientation of the Gabor wavelet, θ

Estimated by CNN

$$G(x,y) = \exp\left(-\frac{X^2 + \boxed{\gamma}^2 Y^2}{2\boxed{\sigma}^2}\right) \times \cos\left(\frac{2\pi}{\boxed{\lambda}} X\right)$$

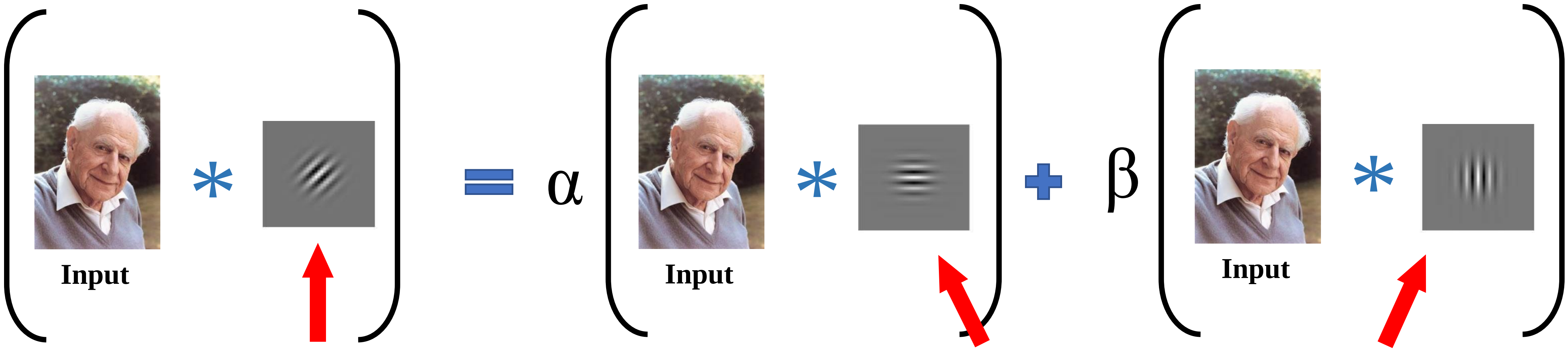
$$X = x\cos\boxed{\theta} + y\sin\theta, Y = -x\sin\theta + y\cos\theta$$

+ Sampling Grid

Not yet estimated

Estimation of Parameter θ

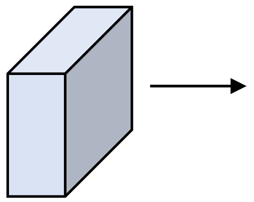
- We address this problem by using a **steering property** that **convolution with Gabor wavelet of any orientation can be represented with a linear combination of a finite set of responses**



Gabor wavelet with arbitrary orientation

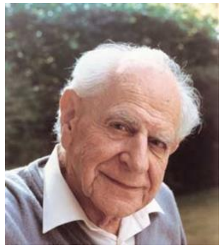
Basis Gabor wavelets

Filtering Part of TGW



Input

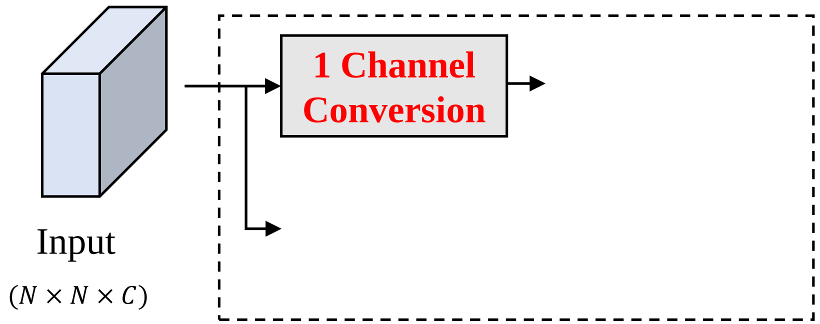
$(N \times N \times C)$

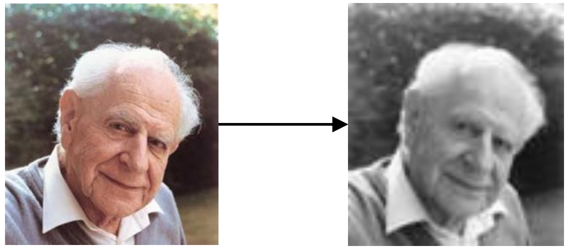


Input Image

Filtering Part of TGW

Filtering by Gabor Filters

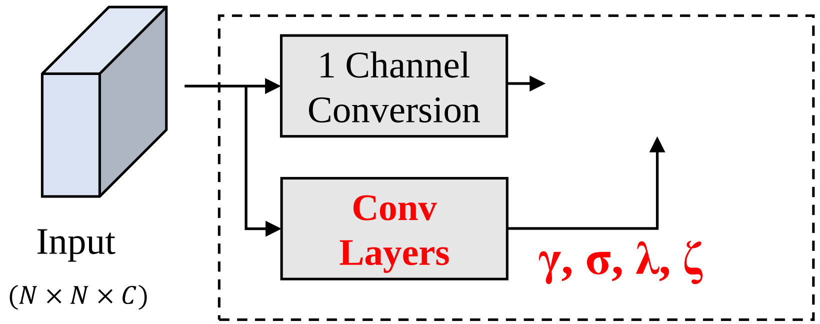


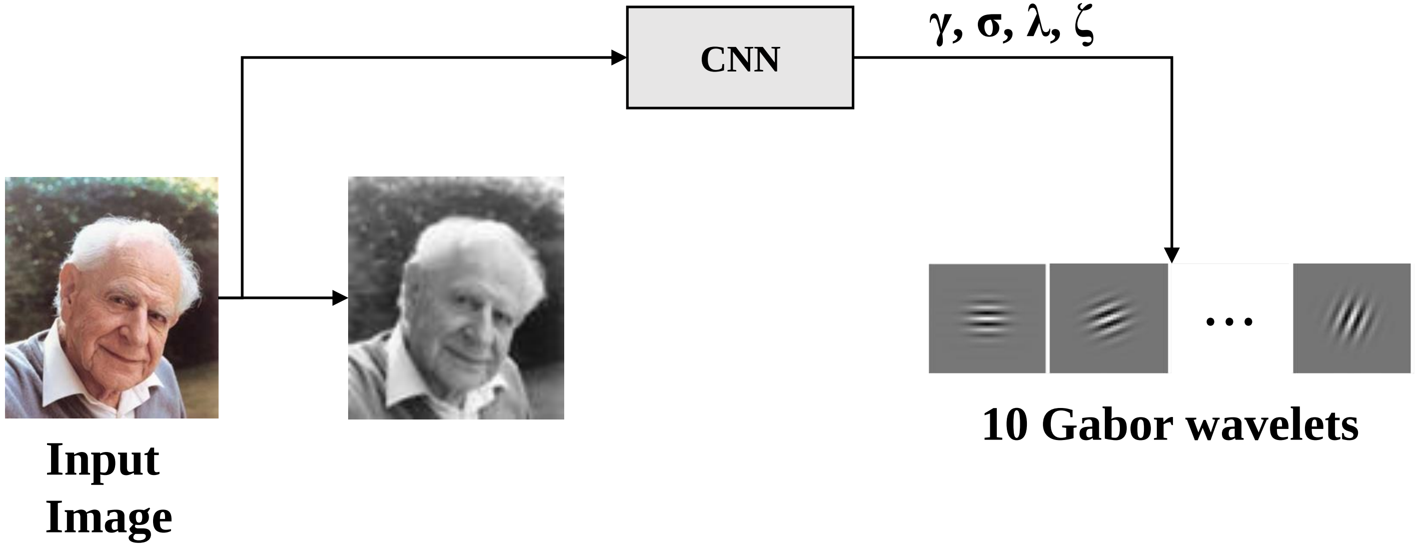


Input Image

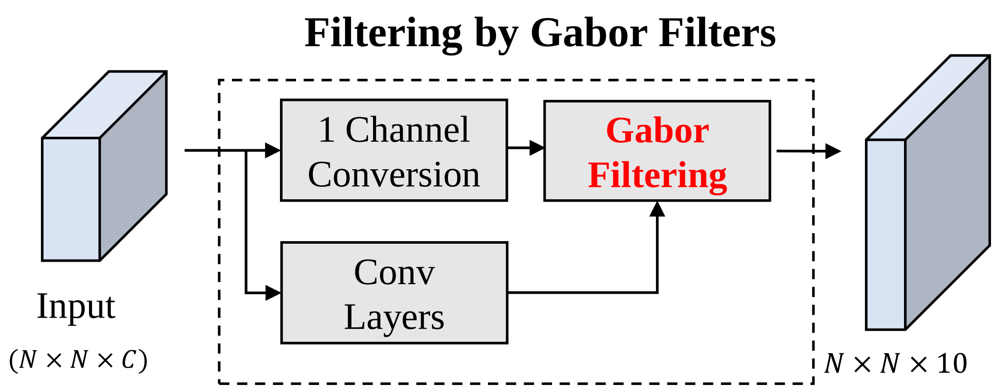
Filtering Part of TGW

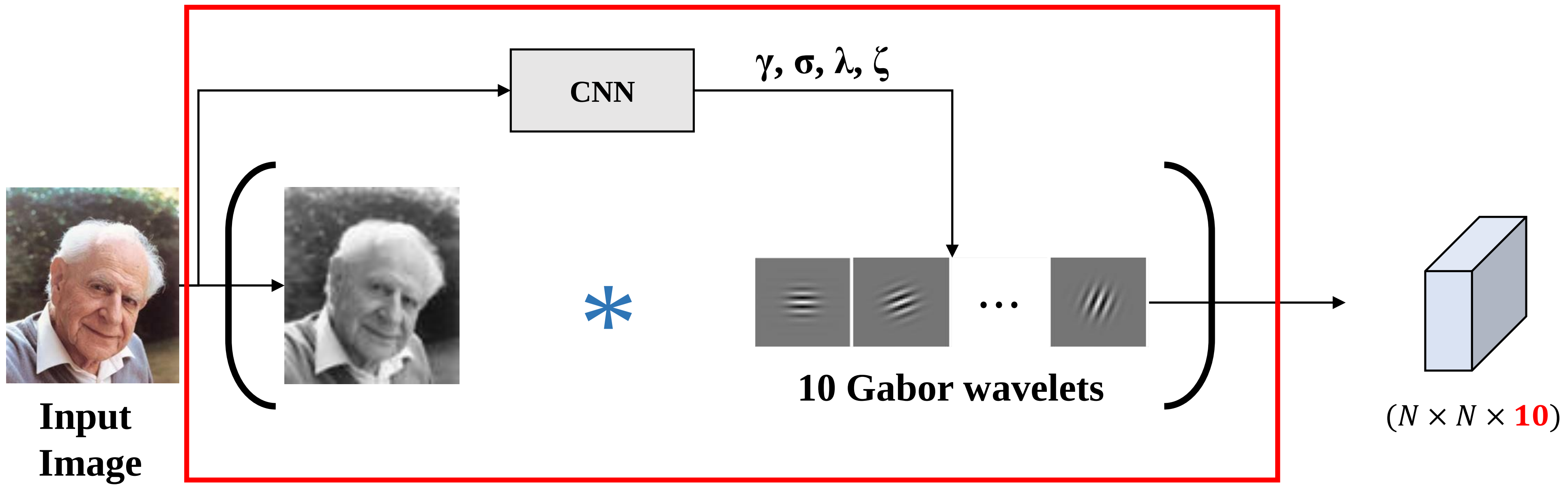
Filtering by Gabor Filters





Filtering Part of TGW



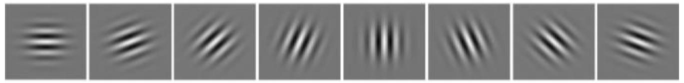


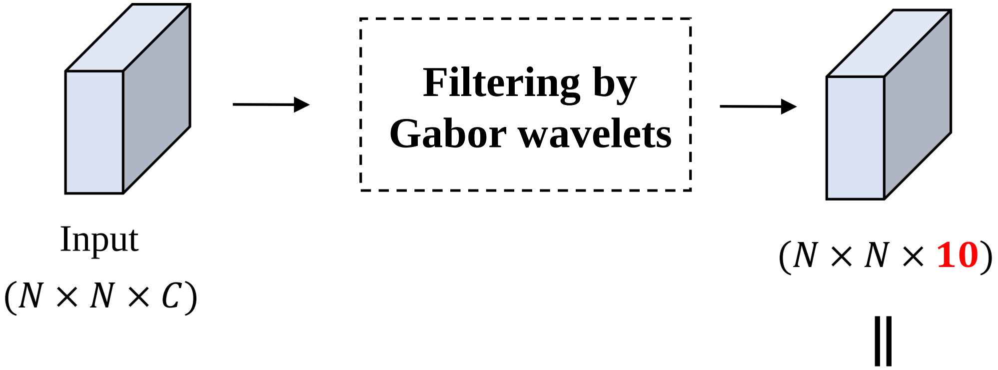
Filtering by Gabor wavelets

Steering Part of TGW

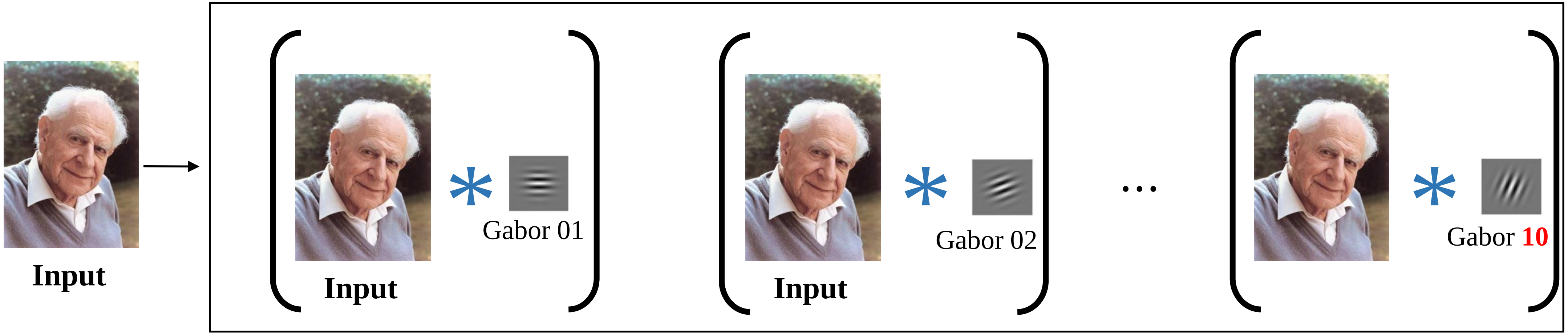
10 Gabor wavelets are used as basis

(θ in [9°, 27°, 45°, 63°, 81°, 99°, 117°, 135°, 153°, and 171°])

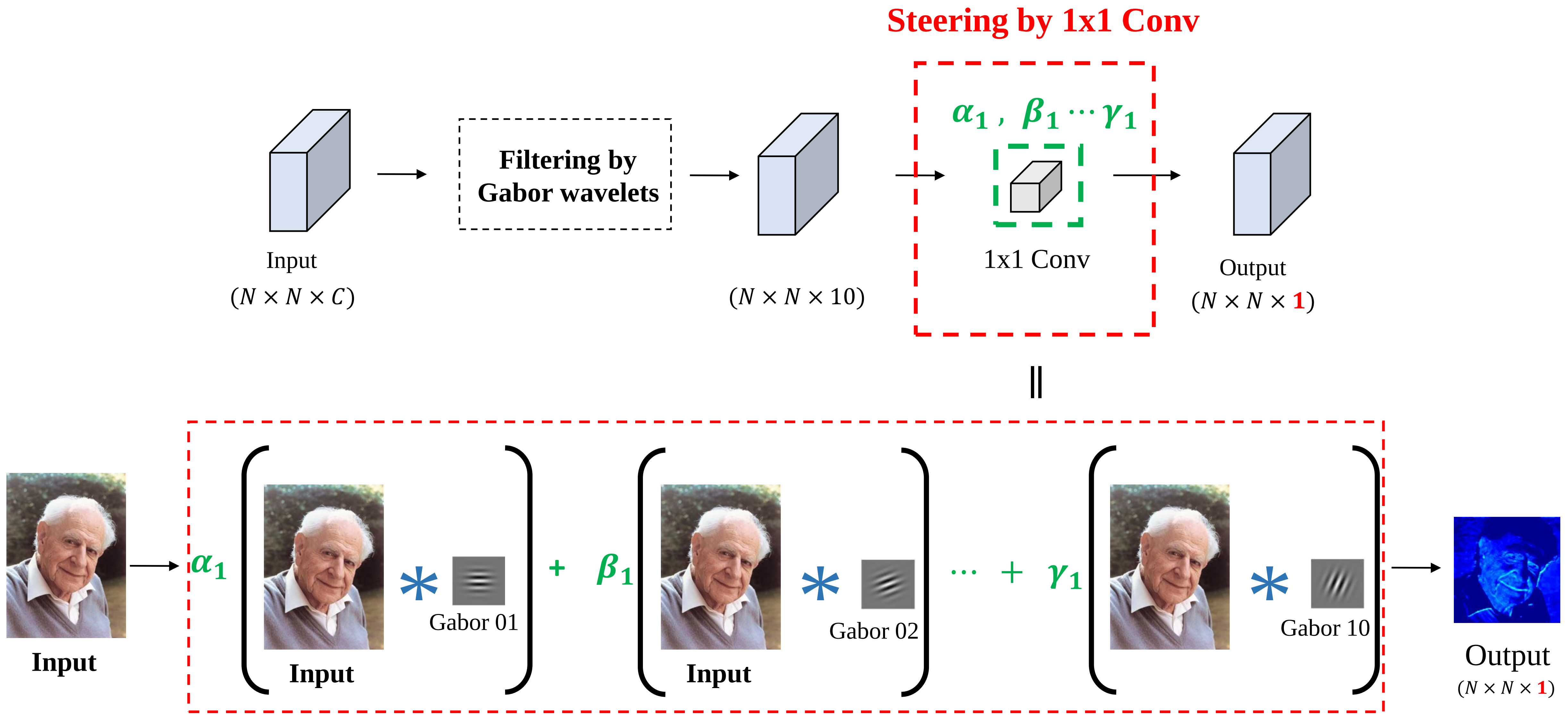




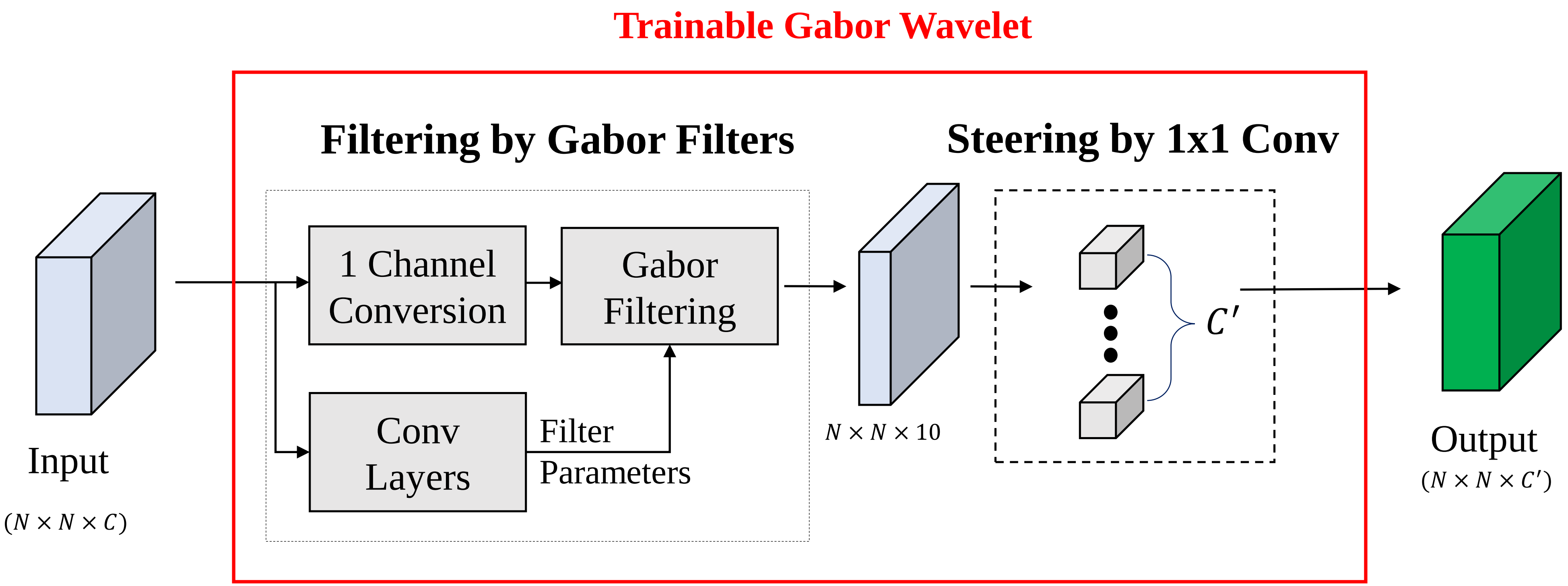
Result of Filtering by Gabor wavelets



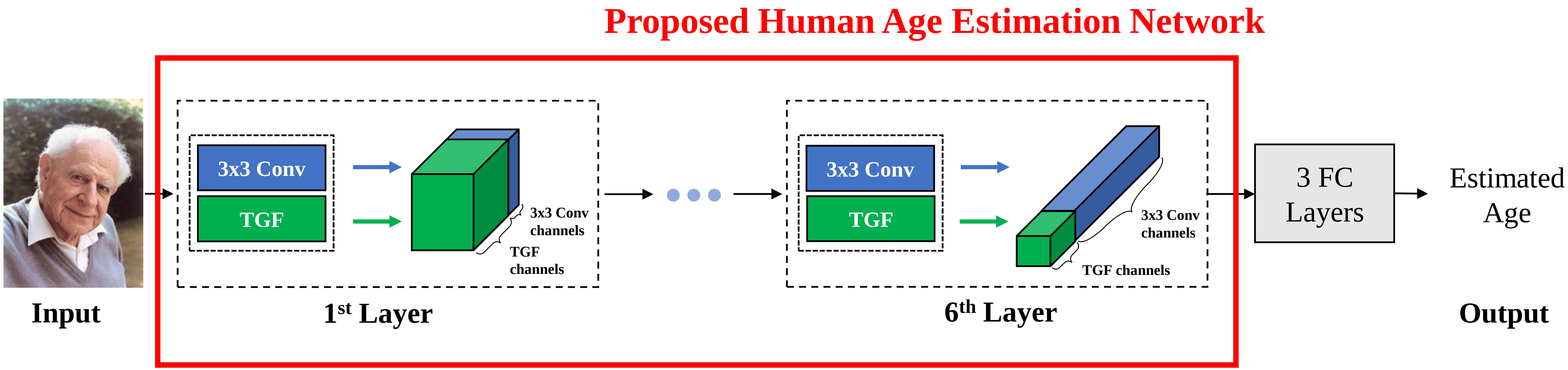
Steering Part of TGW



Trainable Gabor Wavelet (TGW)



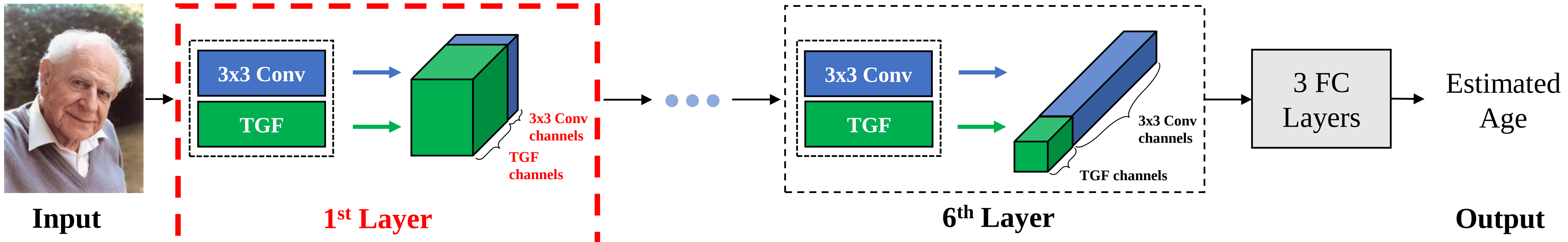
Proposed Human Age Estimation Network



TGF = Trainable Gabor Wavelet
Conv = Convolution
FC = Fully Connected

Assign Different Role to Each Layer

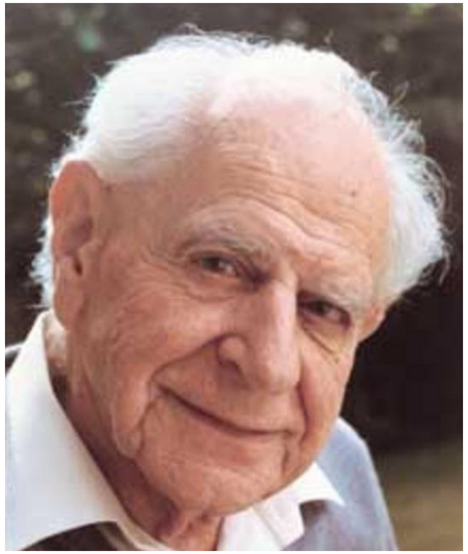
Gabor wavelets act as feature extractor



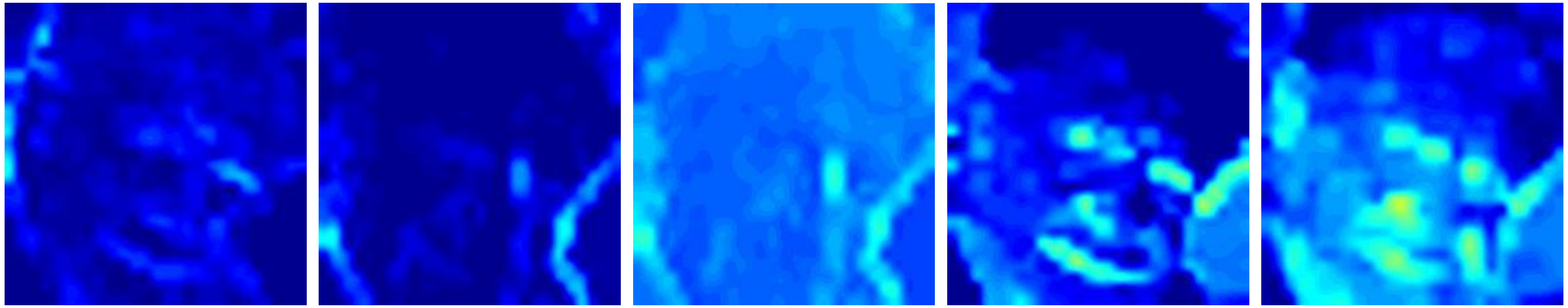
Main Role	**Feature Extraction**	Classification
Dominant Feature	**Features from TGF**	Features from CNN

TGF channels > Conv channels

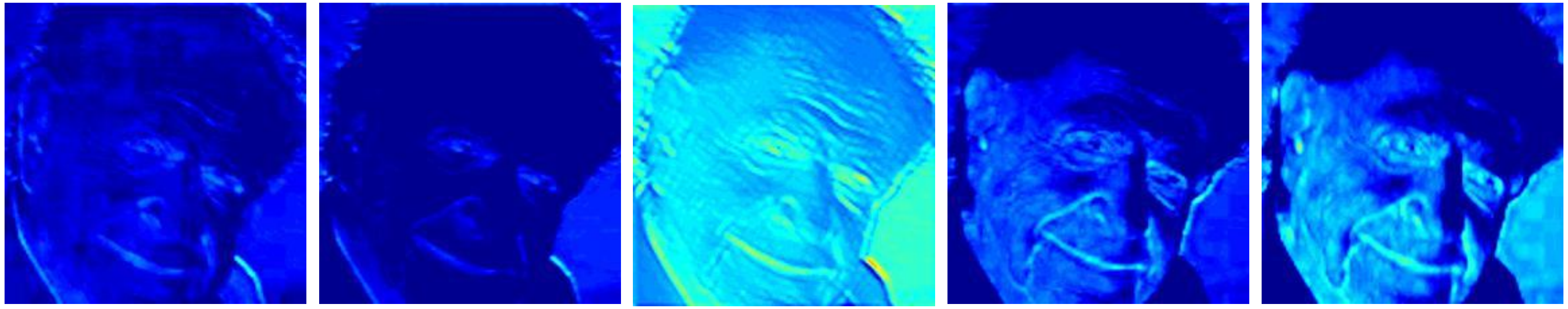
Features from TGW



Input



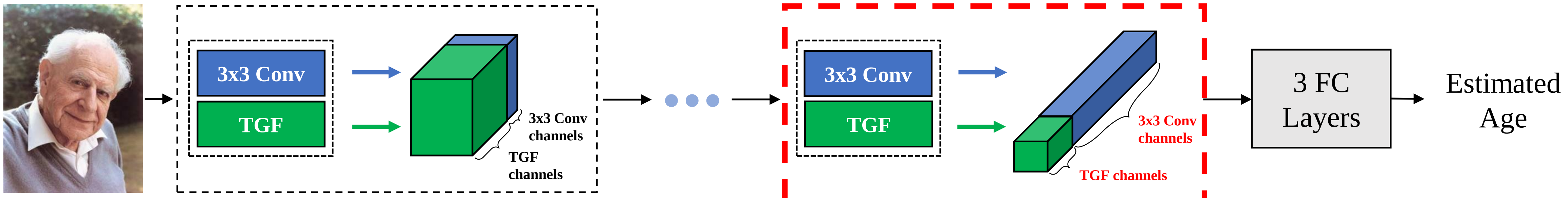
Selected feature map 1st convolution layer of [4]



Selected feature map 1st TGW layer of proposed network

Assign Different Role to Each Layer

Convolution wavelets work as classifier

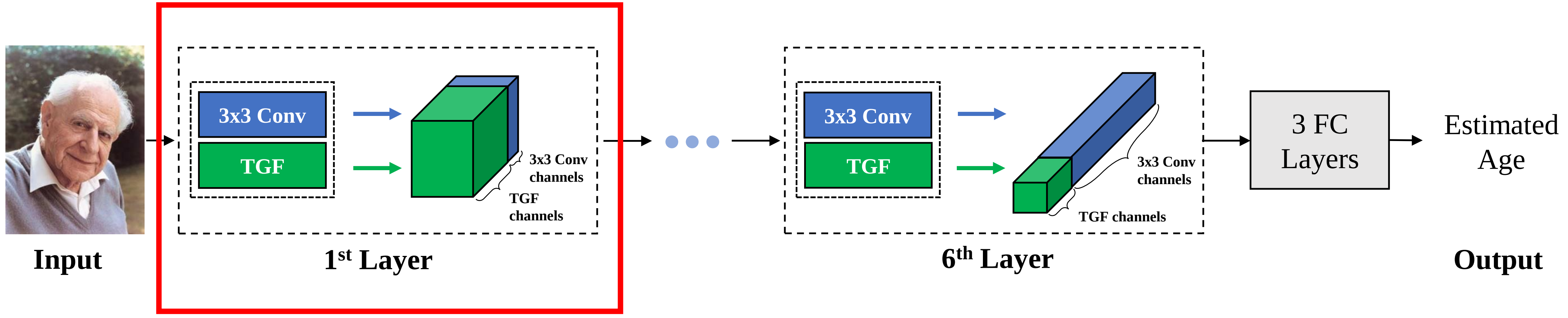


Input | **1st Layer** | **6th Layer** | **Output**

Main Role	Feature Extraction	**Classification**
Dominant Feature	Features from TGF	**Features from CNN**

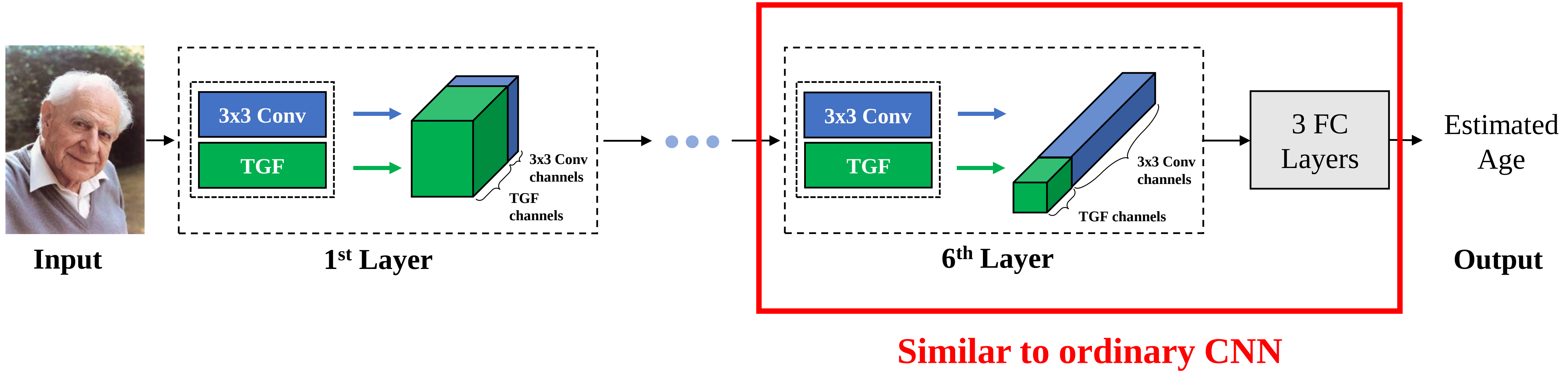
TGF channels < Conv channels

To sum up



Feature extraction with
Trainable Gabor wavelets

To sum up


3x3 Conv
TGF
3x3 Conv channels
TGF channels
Input
1st Layer
3x3 Conv
TGF
3x3 Conv channels
TGF channels
6th Layer
3 FC Layers
Estimated Age
Output
Similar to ordinary CNN

Evaluation

Evaluation Dataset

- Adience Dataset
  - Total 26,580 facial images of 2,284 identities whose ages are labeled in 8 classes
  - Collected from Flicker
  - Performance measure: Classification accuracy (%)

Label (= Class) 0 1 2 3 8

Ablation Study

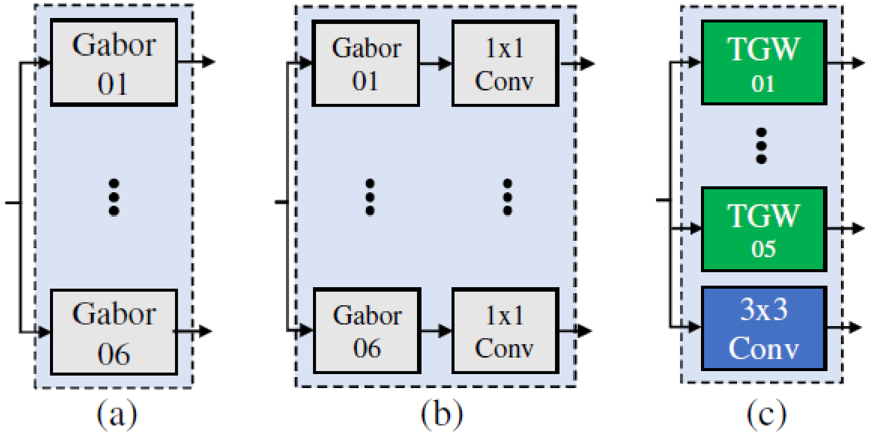


Fig. 4. The first layer in the baseline is replaced with the following structures: (a) the structure for 'M1' and 'M3', (b) the structure for 'M2', and (c) the structure for 'M4' and 'M5'

- M1: six fixed Gabor wavelets w/o a steering block: $\lambda = \gamma = \sigma = 1$ in (1) and pre-defined 16 orientations are used. In this case, 96 orientations are selected to be equally spaced in $[0, \pi)$
- M2: six fixed Gabor wavelets with a steering block: $\lambda = \gamma = \sigma = 1$ in (1) and pre-defined 10 orientations are used.
- M3: six fixed Gabor wavelets w/o a steering block: $\lambda = \lambda_0$, $\gamma = \gamma_0$, $\sigma = \sigma_0$ where $(\lambda_0, \gamma_0, \sigma_0)$ are constants from [5] and pre-defined 16 orientations are used.
- M4: (a) five TGW layers training λ, σ and γ and (b) one 3×3 convolution layer, where $(\lambda_0, \sigma_0, \gamma_0)$ are set as in [5].
- M5: (a) 5 TGW layers training ζ and (b) one 3×3 convolution layer, where $(\lambda_0, \sigma_0, \gamma_0)$ are set as in [5].

Ablation Study

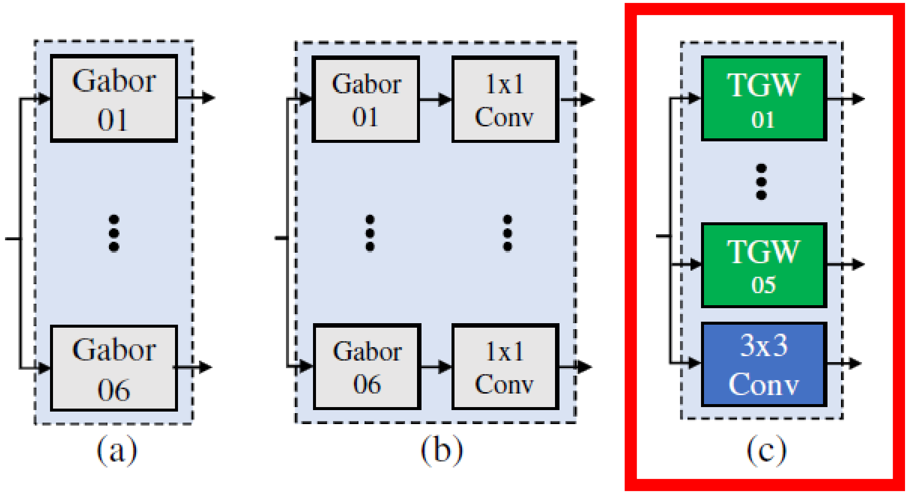
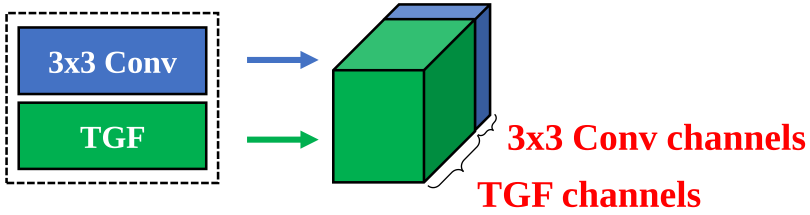


Fig. 4. The first layer in the baseline is replaced with the following structures: (a) the structure for 'M1' and 'M3', (b) the structure for 'M2', and (c) the structure for 'M4' and 'M5'

- M5: (a) 5 TGW layers training ζ and (b) one 3×3 convolution layer, where $(\lambda_0, \sigma_0, \gamma_0)$ are set as in [5].

	M1	M2	M3	M4	M5
Fixed filters	6	6	6	-	-
TGW layers	-	-	-	5	5
CONV layers	-	-	-	1	1
Steering	-	O	-	O	O
Initialization	with 1s	with 1s	[5]	-	-
Accuracy (%)	47.61	48.87	49.93	49.70	**51.00**

Settings of The Proposed Method



Layer	TGW channels	3x3 Conv channels
1	180	40
2	155	80
3	130	160
4	105	240
5	80	320
6	55	400

Age Estimation Results

Table 3. Age classification result on Adience dataset.

Method	Accuracy (%)	Parameters
LBP [23]	41.1	-
LDP [18]	48.5	-
LDN [18]	51.4	-
Best from [4]	50.7	11M
Extension of [4]	53.4	18M
Resnet-50 [17, 20]	52.2	25M
Cascaded DCNN [19]	52.9	40M
PTP [18]	53.3	-
WC-CNN [24]	54.3	$>$ 62M
Ours (Proposed)	**54.4**	**18M**

[4]: Levi, Gil, and Tal Hassner. "Age and gender classification using convolutional neural networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015.

Conclusion

Conclusion

- We have proposed the Trainable Gabor Wavelet that can incorporate Gabor wavelets into CNN in trainable manner.

- We also built an age estimation network with the proposed Trainable Gabor Wavelet.

- Proposed age estimation network showed better results with parameter efficiency.

Limitations and Future works

- TGF is not tested on the general-purpose dataset like ImageNet

  → Experiments will be done in these tasks.

Q & A

Thank You