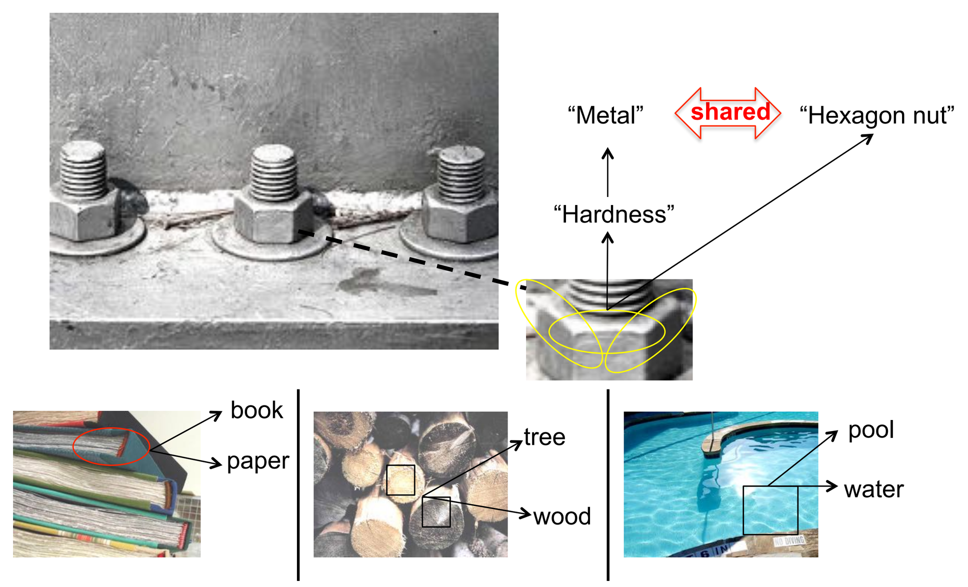
Learning Deep Representations of Objects and Materials for Material Recognition

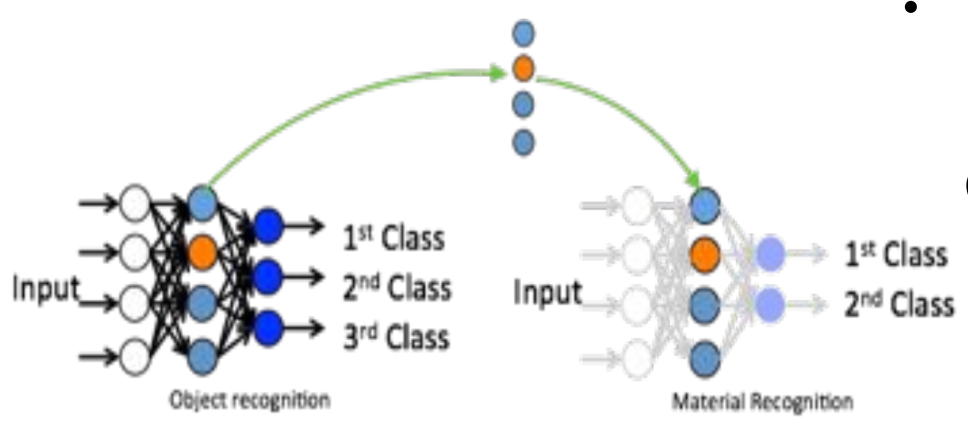
Xing Liu, Mete Ozay, Yan Zhang, Takayuki Okatani
Graduate School of Information Sciences, Tohoku University

- We study how representations learned for object recognition can be used for material recognition.
- We compare our CNN models with human vision systems in terms of recognition accuracy of material category using natural images and their deformed versions.

Motivation

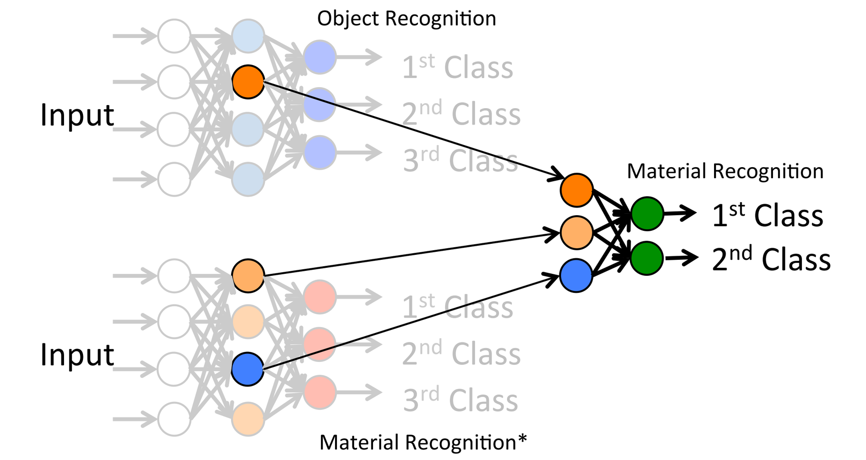


- Despite the recent success in the application of convolutional neural network(CNNs) to object category recognition, there are many open questions regarding material category recognition, i.e., recognizing material of an object from its single image. In this study, we analyze and explore how to utilize CNN features learned for object recognition for material recognition.



- Toward this end, we conducted the following two experiments:

(i) Transfer Learning:
We first learn the object features using a CNN and then transfer the learned representations for material recognition task.



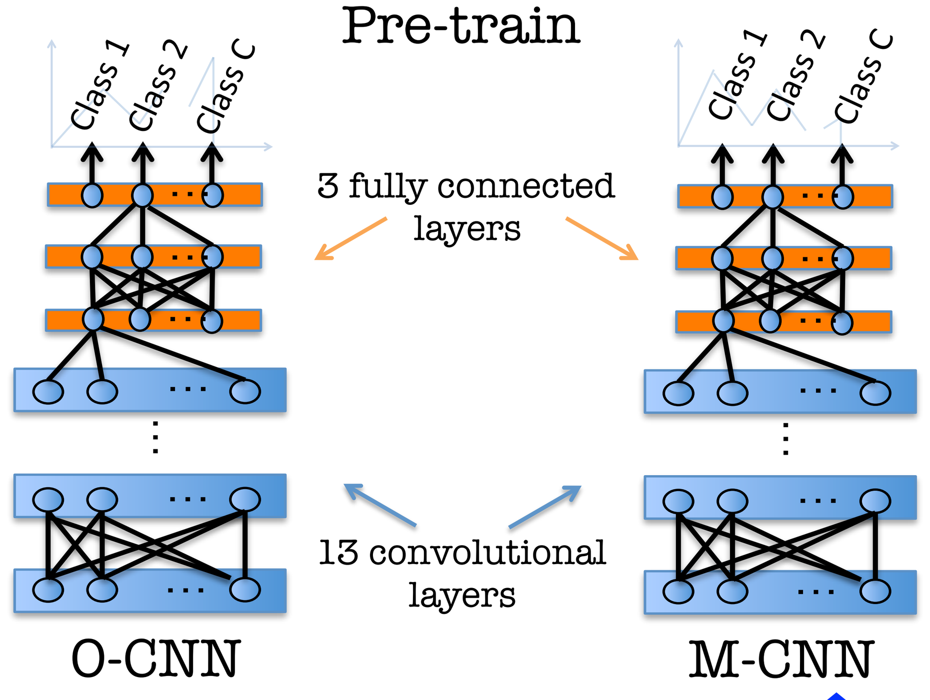
(ii) Feature Selection and Integration:
We learn object features and material features separately using two different CNNs. Then, we select and integrate these features in order to recognize material category as accurately as possible.

- We report the following two results:

(1) Accuracy of the above two methods on material category recognition

(2) Comparison of human vision and CNNs for material recognition using natural images and their deformed versions.

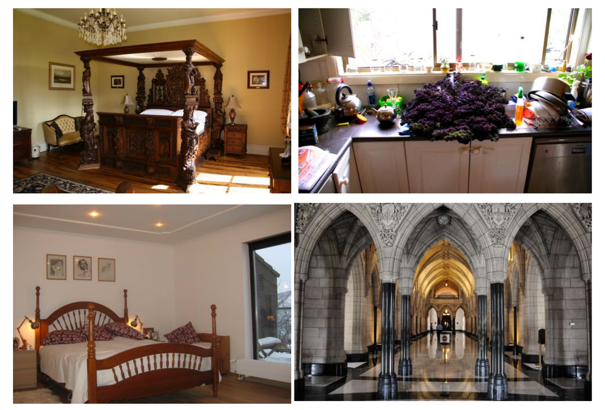
Database:



ImageNet:
(Object Database)

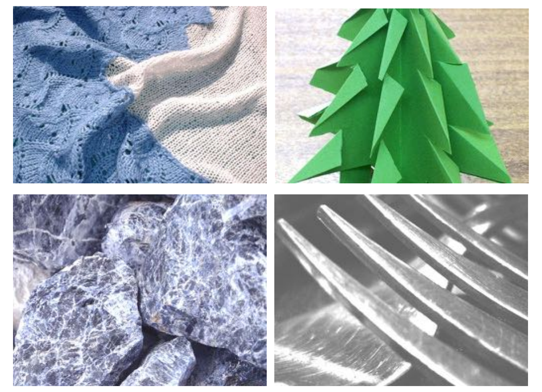
categories: 1000
images: 1 million

Minc:
(Material-scene Database)



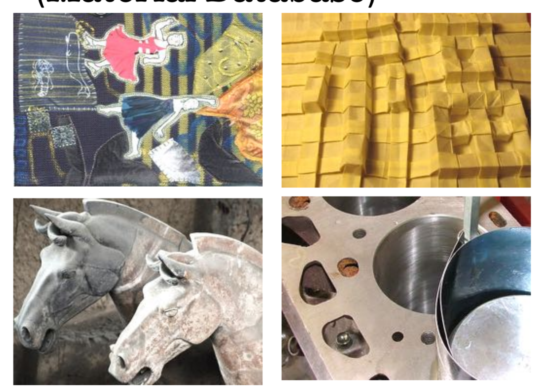
categories: 23
images: 3 million

FMD:
(Material Database)



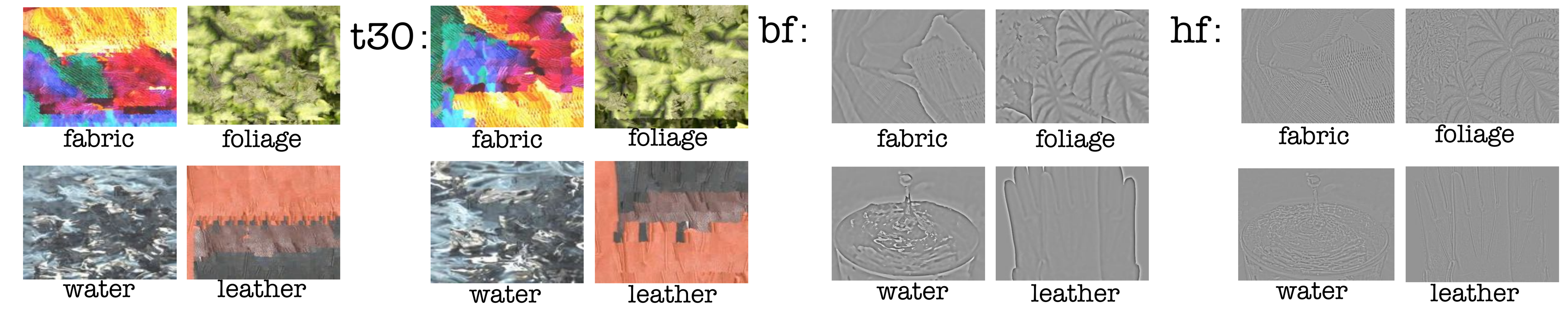
categories: 10
images: 1,000

EFMD(Extend-FMD):
(Material Database)



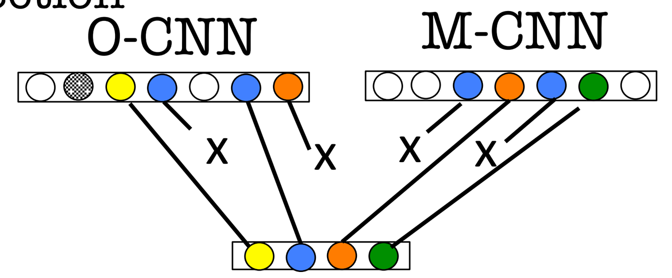
categories: 10
images: 10,000

Deformed_FMD:



Approach & Results

Feature Selection

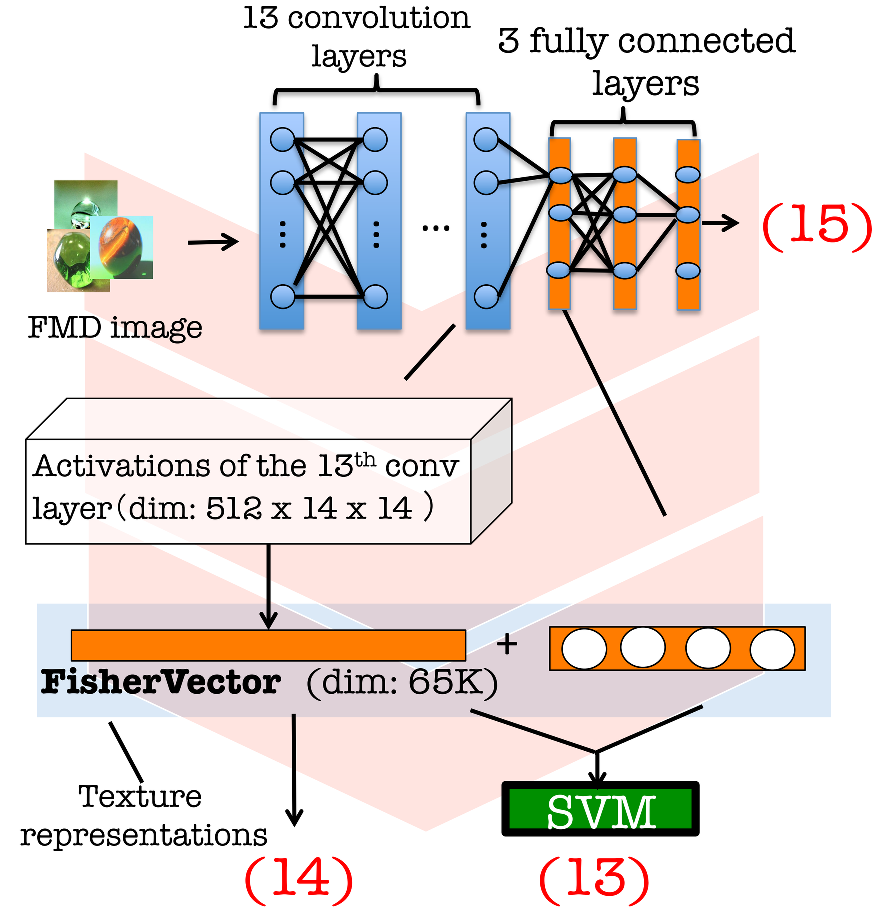


$$\mathcal{H}_{(Y)} = -\sum_{(x_c^i, Y^i)\in\mathcal{T}} \omega_i \cdot P(x_c^i, Y^i|x_{c,j}) log P(x_c^i, Y^i|x_{c,j})$$

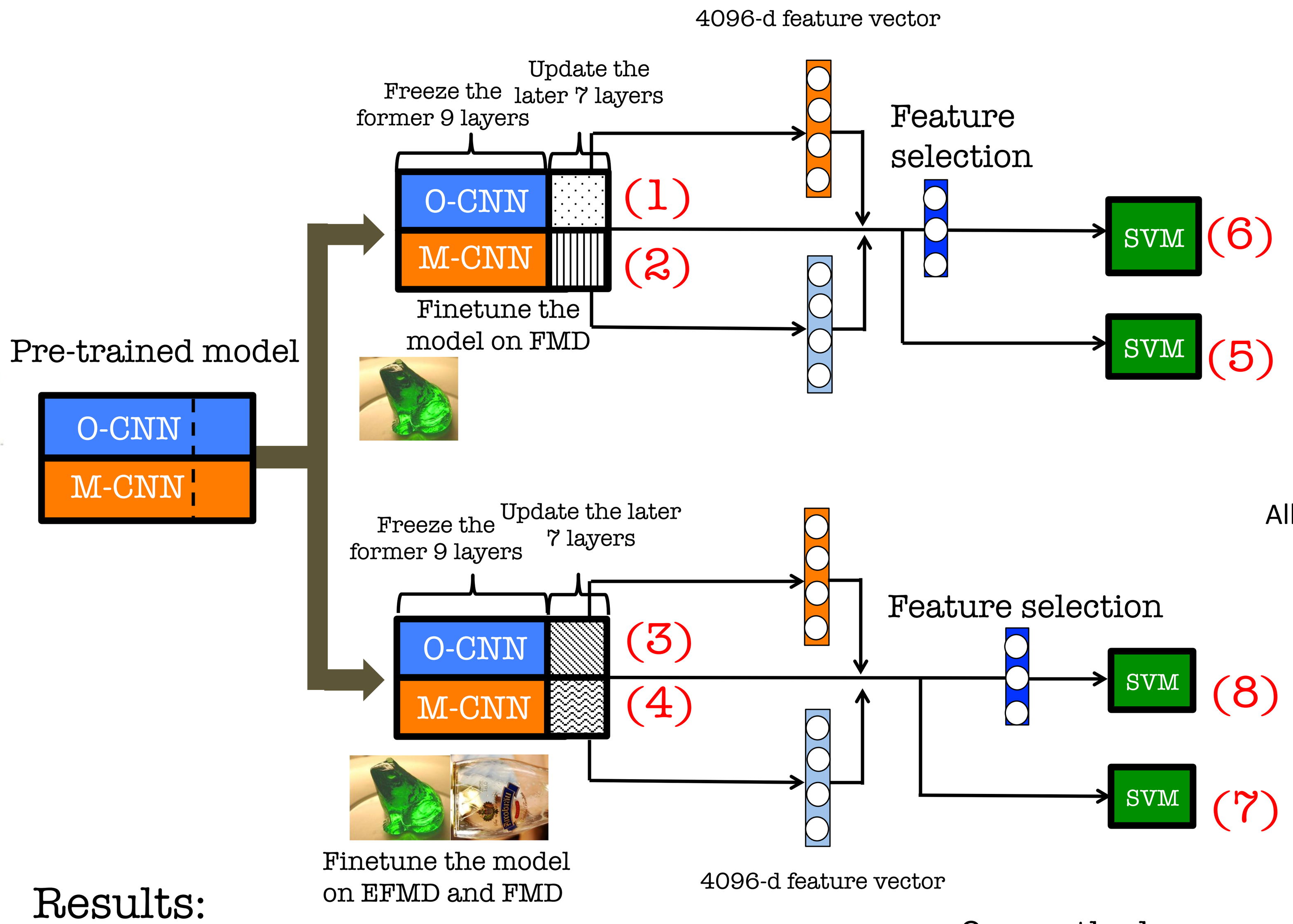
$$\omega_k = \omega_k * (1 + \mathcal{H}_j(Y)^{-1})$$

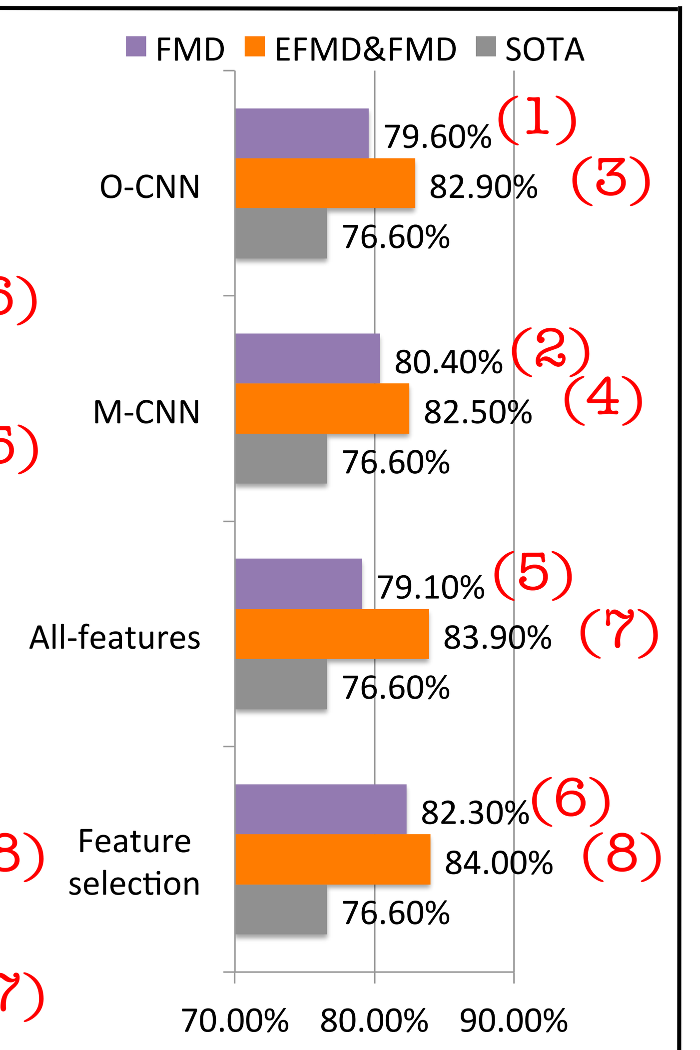
$$x_{c,\delta}^* = argmin \sum_{(x_c^i, Y^i)\in\mathcal{T}} \mathcal{H}(Y)$$

Related Work **[Cimpoi et al. 2015]**



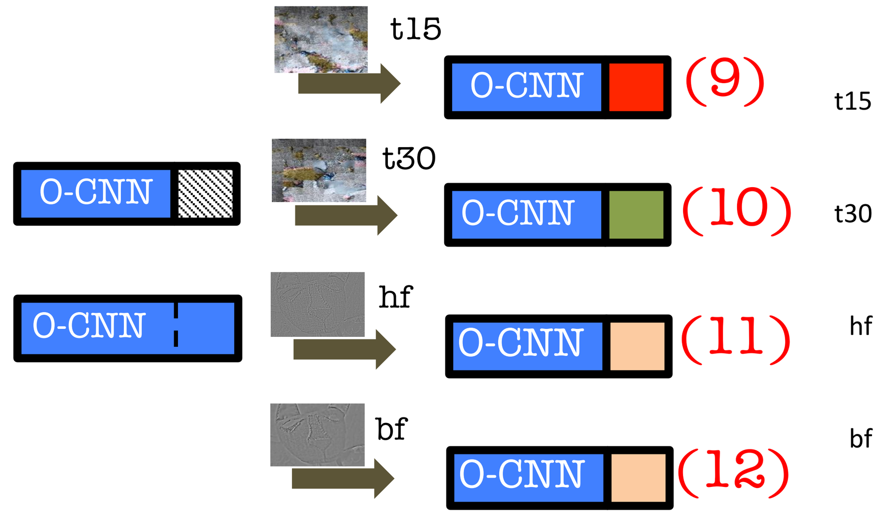
	VGG-FC	VGG-FV	VGG-FC+FV
VGG16	(15) $70.3\%_{\pm 1.8}$	(14) $73.5\%_{\pm 2.0}$	(13) $76.6\%_{\pm 1.9}$
VGG19	$77.4\%_{\pm 1.8}$	$79.8\%_{\pm 1.8}$	**$82.4\%_{\pm 1.5}$**

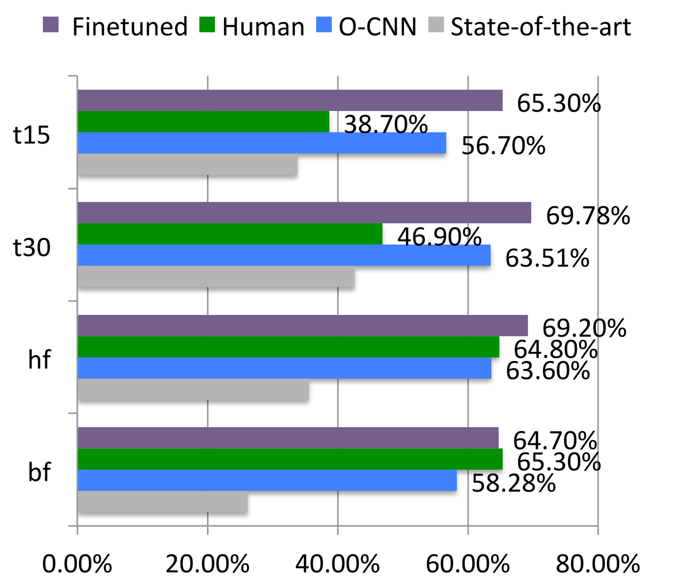




Results:

- CNNs trained in an end-to-end, transfer learning setting performs very well, especially when they are trained on a large training dataset.
- Our method proposed for selection/integration of material and object features learned by CNNs is also effective, particularly when a small training dataset is used.





Conclusion:

- Humans perform better than CNNs in material recognition, probably because they use global information more effectively.
- CNNs are better at utilizing more local information to recognize materials.