

#### How Useful Are the Machine-Generated Interpretations to General Users? A Human Evaluation on Guessing the Incorrectly Predicted Labels

Hua Shen, Ting-Hao (Kenneth) Huang

College of Information Sciences and Technology The Pennsylvania State University {huashen218, txh710}@psu.edu





### **Are Machine-Generated Interpretation Useful?**

# The model misidentified this image:



Correct Label: Movie Theater

# **Are Machine-Generated Interpretation Useful?**

# The model misidentified this image:





### **Are Machine-Generated Interpretation Useful?**



# **Thesis Statement**

Displaying the visual interpretations to general users did not increase, but rather decreased, the average accuracy on guessing incorrectly predicted labels by roughly 10%.



#### The model misidentified





#### The model misidentified



















#### **Results**

	C1	C2	C3	C4	C5	Overall
Int	0.57	0.74	0.66	0.41	0.67	0.63
No-Int	0.52	0.71	**0.84	*0.59	0.77	**0.73
#images	44	20	112	18	6	200

Table: Average Human Accuracy on Inferring ModelMisclassification (non-overlap users).

#### Conclusion

[No-Interpretation] condition > [Interpretation] condition

(statistically significant)

10% Accuracy Drop

# What does this mean?

Be cautious when displaying machine-generated interpretations to explain models' mistakes, because it is not always helpful for understanding by general users.

#### **Model Error Categories for Fine-Grained Analysis**



### **Model Error Categories for Fine-Grained Analysis**



# Thank you!





Hua Shen (huashen218@psu.edu)

#### Ting-Hao(Kenneth) Huang (txh710@psu.edu)



#### GitHub: https://github.com/huashen218/GuessWrongLabel

huashen218 / GuessWrongLabe	l.	⊙ Unwatch + 2	☆ Star 2 양 For		
<> Code ① Issues ① Pull request	Actions III Projects III Wiki () Security	🗠 Insights 🕸 Settings			
같 master → 같 1 branch ⓒ0 tags	Go to file Add	file - 🛓 Code - About			
buashen218 replace arxive link	afd67ee on	No description	on, website, or topics		
Interpretation_Interface	change ui name	last month 🛛 Readme	🛱 Readme		
No_Interpretation_Interface	change ui name	last month			
condition1	add ui demo	last month Releases	Releases		
condition2	add interface codes	last month Create a new rel	No releases published Create a new release		
machine_generated_interpretation	add ui demo	last month			
DS_Store	replace arxive link	last month Packages	Packages		
README.md	replace arxive link	last month No packages pul	ast month No packages published		
UI_Interface_Demo.png	add ui png	Publish your first package			
UI_Interface_Long.png	add ui png	last month			
README.md Human Evaluation	on Interpretation by GuessWro	HTML 95.19     Shell 0.2%	• Python 4.7%		
GuessWrongLabel User Interface (I	II) was developed to support the paper:				
<ul> <li>How Useful Are the Machine-G Wrongly Predicted Labels. Hua Computation and Crowdsourcir</li> </ul>	enerated Interpretations to General Users?A Human Evaluatic Shen, Ting-Hao (Kenneth) Huang. The 8th AAAI Conference g (HCOMP), 2020.	n on Guessing the on Human			

in this report the six examples.

- UI for [No-Interpretation] condition is in "/No\_Interpretation\_Interface/" with original images in "/condition1/";
- UI for [Interpretation] condition is in "/Interpretation\_Interface/" with images and saliency maps in "/condition2/".