Building Usage Profiles Using Deep Neural Nets

Domenic Curro, Konstantinos G. Derpanis, Andriy V. Miranskyy
Department of Computer Science, Ryerson University, Toronto, Canada
{d2curro, kosta, avm}@scs.ryerson.ca

Motivation

Objective

Build a software usage profile from tutorial videos hosted online

Challenges

Individual user actions are a latent variable within video frames

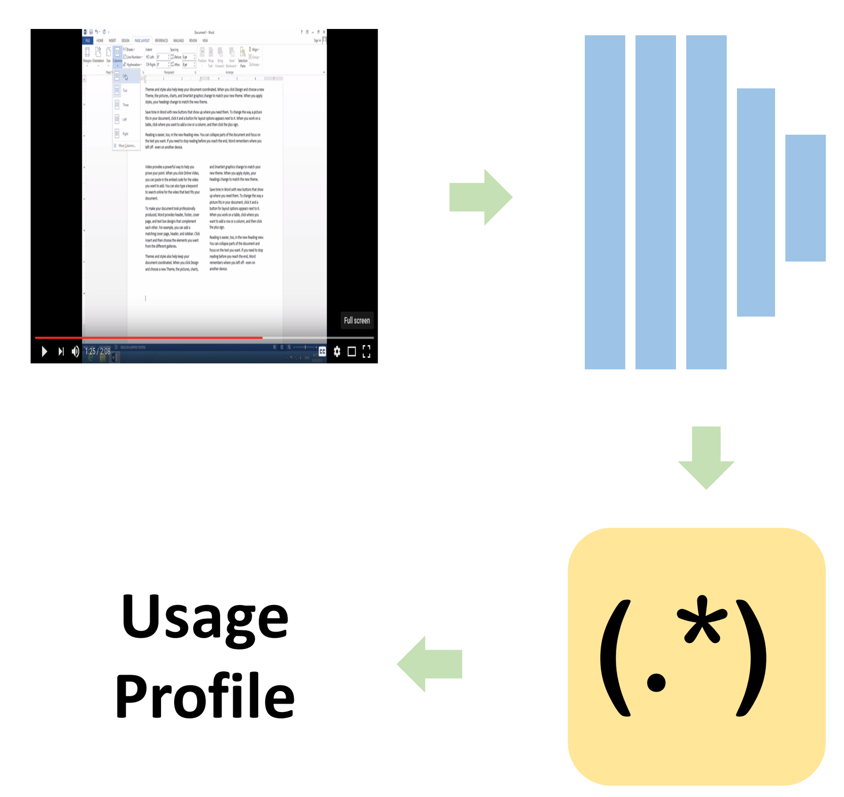
Prior Work

Video recordings have been used in ethnographic studies, for user experience [1]. Automatic tools exist that can locate specified images on screen to some threshold [2]. There are suggestions that DCNNs can be used in source code analysis [3]. To the best of our knowledge, nobody has considered extracting user actions from UI-based images

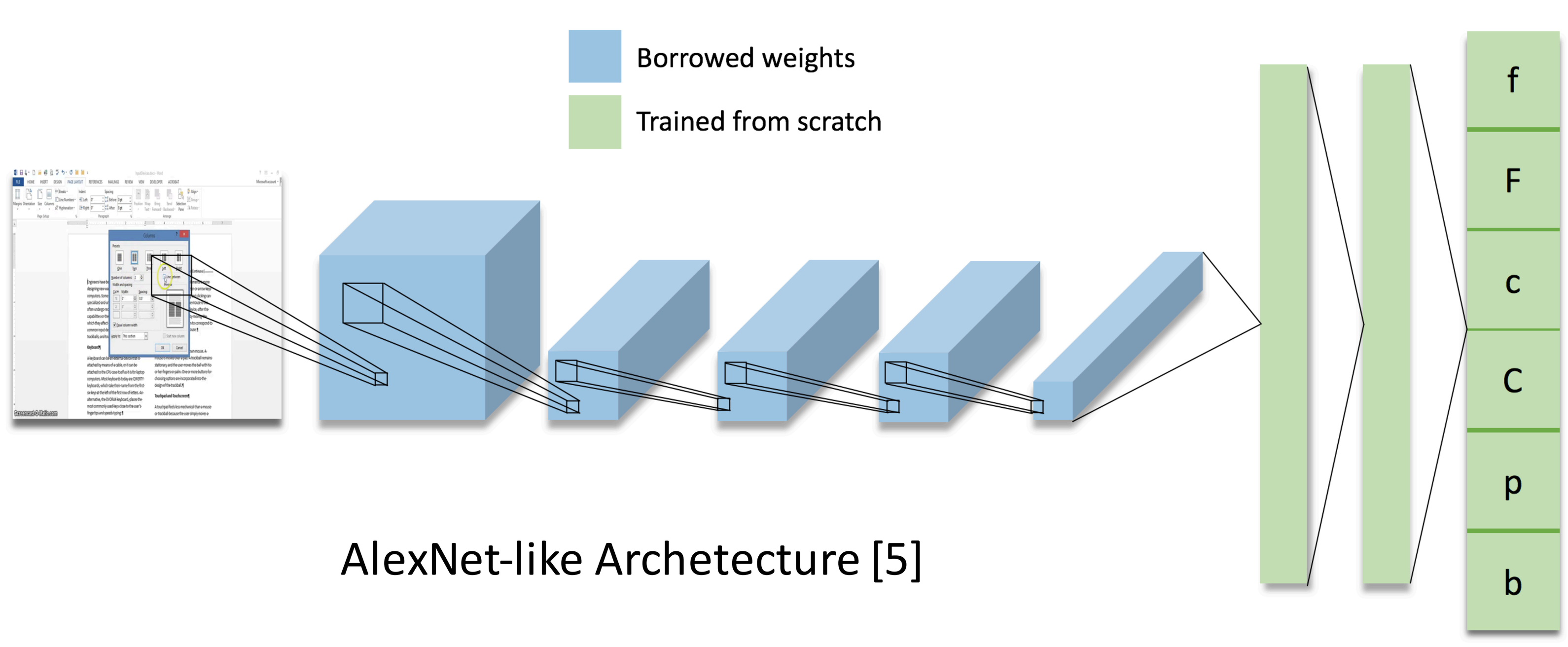
Contribution

Automatic proof-of-concept pipeline for generating usage profiles from raw video data

Pipeline



Action Sequence



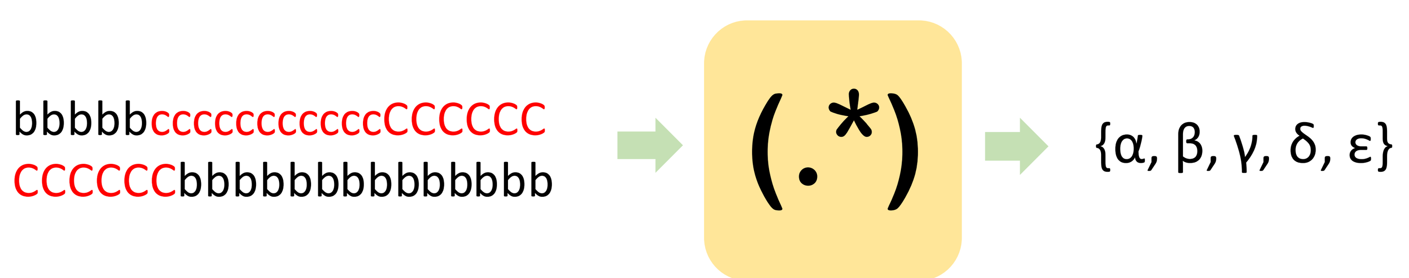
AlexNet-like Archetecture [5]

- Pre-trained network with ImageNet [4] weights
- Sufficiently trained with very few examples per class
- Invariant to image noise, screen capture artifacts, various Word versions, screen resolutions, system fonts, system colours, and mouse occlusions

User Profile	Data Count	Apparent User Action
f	118	The font menu is open
F	46	The default font window is open
c	381	The columns drop down is open
C	396	The columns window is open
p	293	The page number drop down is open
b	39.5k	None of the above user actions is occurring

Evaluating a video, a sequence of actions is collected

User Profile Prediction

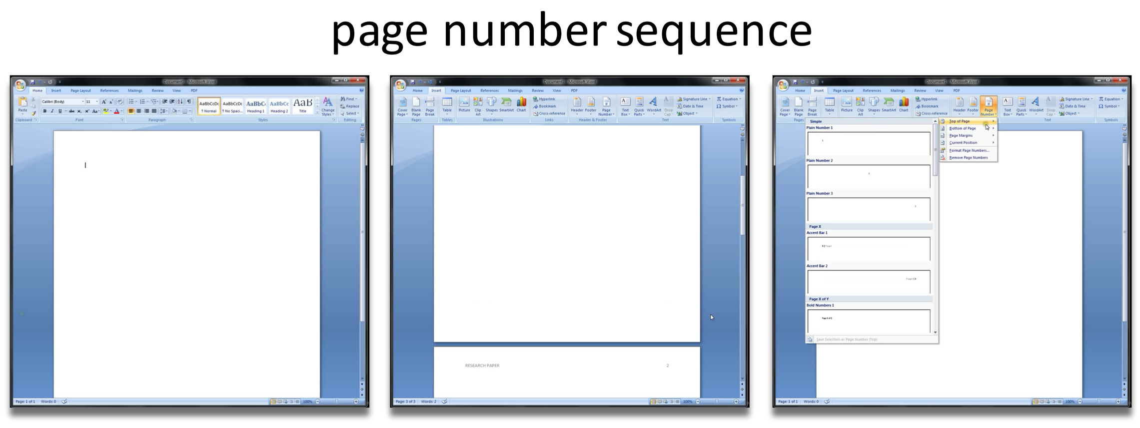


A sequence of user actions are matched against a series of regular expressions, selecting the appropriate usage profile

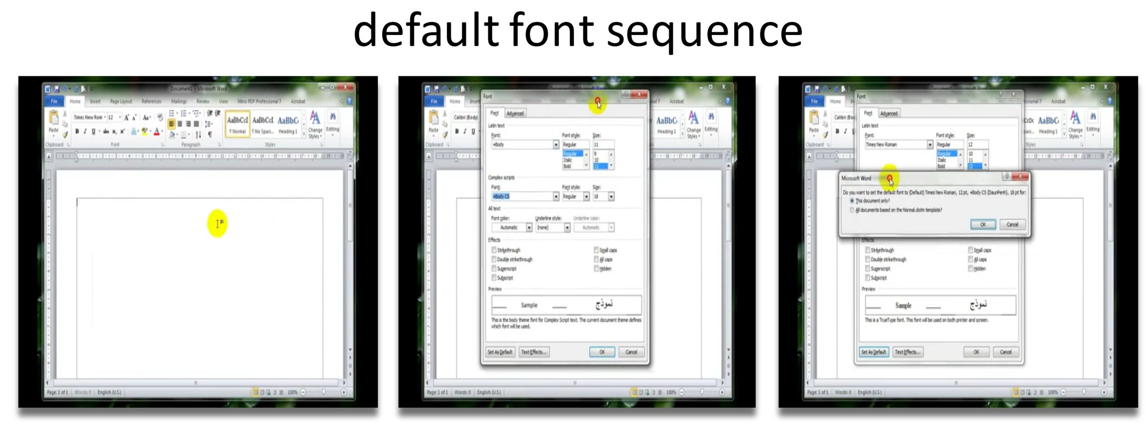
User Profile	Regular Expression	The user
α	`f{r,}`	...changed font via the font menu
β	`f{r,}F{r,}f{0,r}`	...changed their default font
γ	`c{r,}`	...changed the column count via the drop down menu
δ	`c{r,}[^cC]{0,r}C{r,}`	...changed the column count via the column window
ε	`p{r,}`	...changed page numbering via the page number dropdown

Results

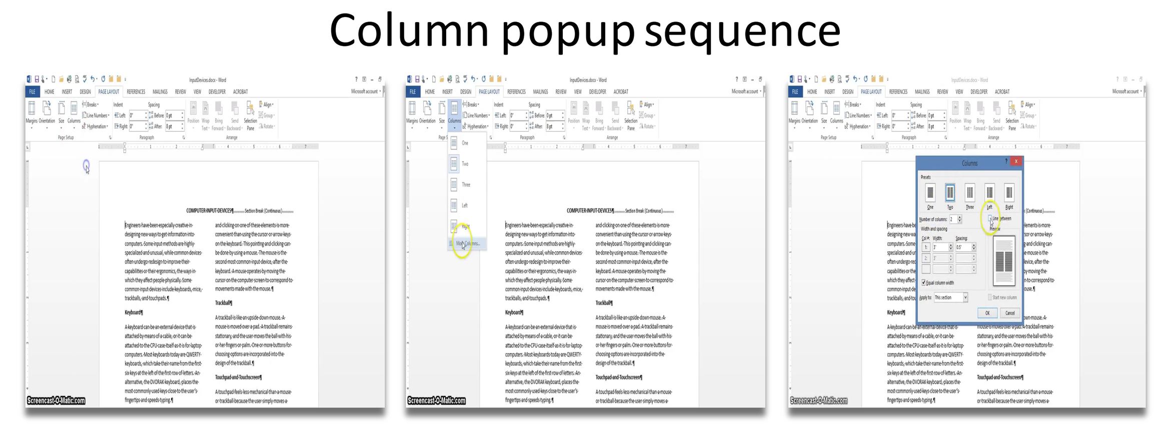
page number sequence



default font sequence



Column popup sequence



Action Prediction Confusion Matrix

	b	f	F	c	C	p	Recall
b	38852	49	25	210	170	198	98.35%
f	27	86	2	0	3	0	72.88%
F	8	4	33	0	1	0	71.74%
c	34	0	0	347	0	0	91.08%
C	93	0	1	2	300	0	75.76%
p	39	0	0	0	1	253	86.35%
	99.49%	61.87%	54.10%	62.08%	63.16%	56.10%	precision
	98.91%	66.93%	61.68%	73.83%	68.89%	68.01%	F1-score

Individual Actions Mean F1-score (without **b**): 67.87%

Profile Prediction Confusion Matrix

	α	β	γ	δ	η	Recall
α	4	1	1	0	0	66.67%
β	2	15	0	0	0	88.24%
γ	1	0	75	2	2	93.75%
δ	0	0	4	54	0	93.10%
η	0	0	1	0	74	98.67
	57.14%	93.75%	92.59%	96.43%	97.37%	precision
	61.54%	90.91%	93.17%	94.74%	98.01%	F1-score
	80.16%	95.10%	97.25%	99.82%	99.79%	AP

Usage Profile Mean Average Precision: 94.42%

References

1. D. Socha, R. Adams, K. Franznick, and et al. Wide-field ethnography: Studying software engineering in 2025 and beyond. In *38th Int. Conf. on Software Engineering Companion*, pages 797–802, 2016.
2. T. Yeh, T.-H. Chang, and R. C. Miller. Sikuli: using GUI screenshots for search and automation. In *22nd Annual Symp. on User interface software and technology*, pages 183–192, 2009.
3. M. White. Deep representations for software engineering. In *37th Int. Conf. on Software Engineering*, volume 2, pages 781– 783, 2015.
4. ImageNet large scale visual recognition competition 2012 (ilsvrc2012), 2012. http://image- net.org/challenges/LSVRC/ 2012/browse-synsets.
5. A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.