Supplementary Material
Can Deep Learning Recognize Subtle Human Activities?

Vincent Jacquot[1], Zhuofan Ying[2], and Gabriel Kreiman[3,4]

jacquot.vinc@gmail.com,zuofanying@gmail.com,gabriel.kreiman@tch.harvard.edu
[1]Ecole Polytechnique Federale de Lausanne
[2]University of Science and Technology of China
[3]Children's Hospital, Harvard Medical School
[4]Center for Brains, Minds, and Machine

S1. Supplementary figures

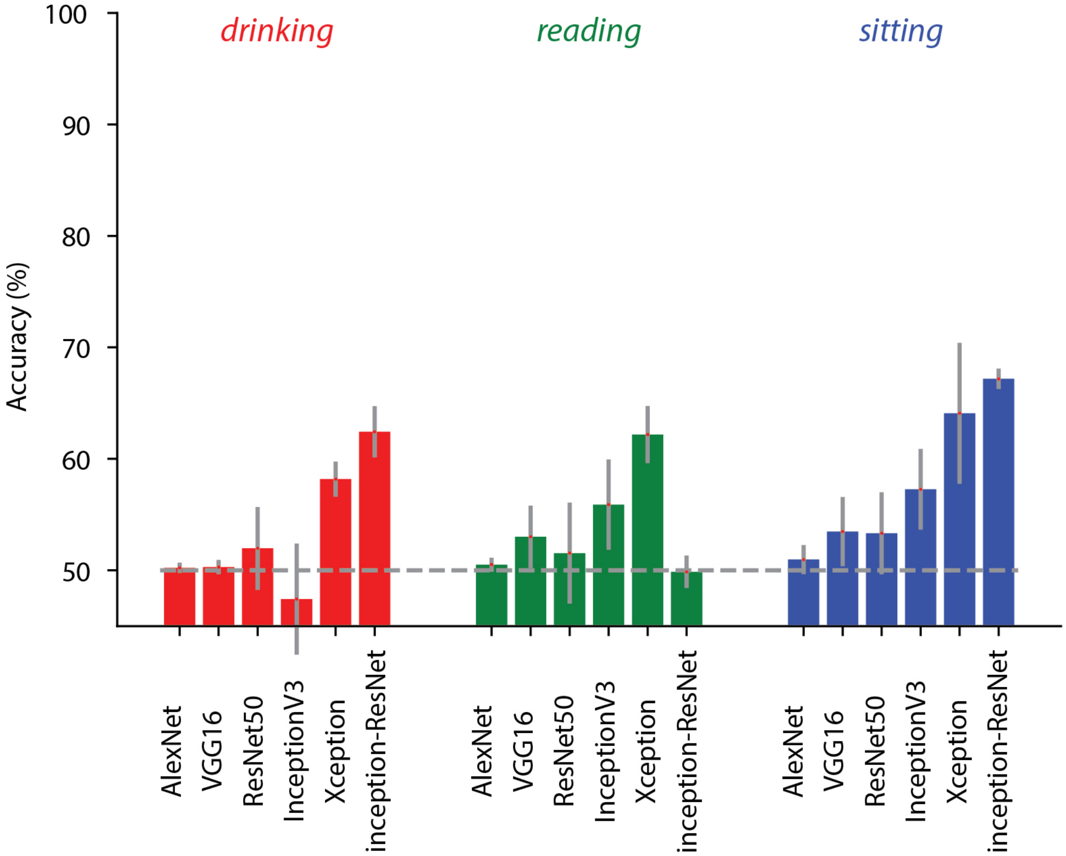

Figure S1. **Performance of deep convolutional neural network models in action recognition using RGB images**. This figure follows the conventions and format of **Figure 5** in the main text. Here we present results using RGB images. Test performance for each fine-tuned model is shown ($mean \pm SD$). The model with best accuracy on the validation set was retained to be applied on the test set.

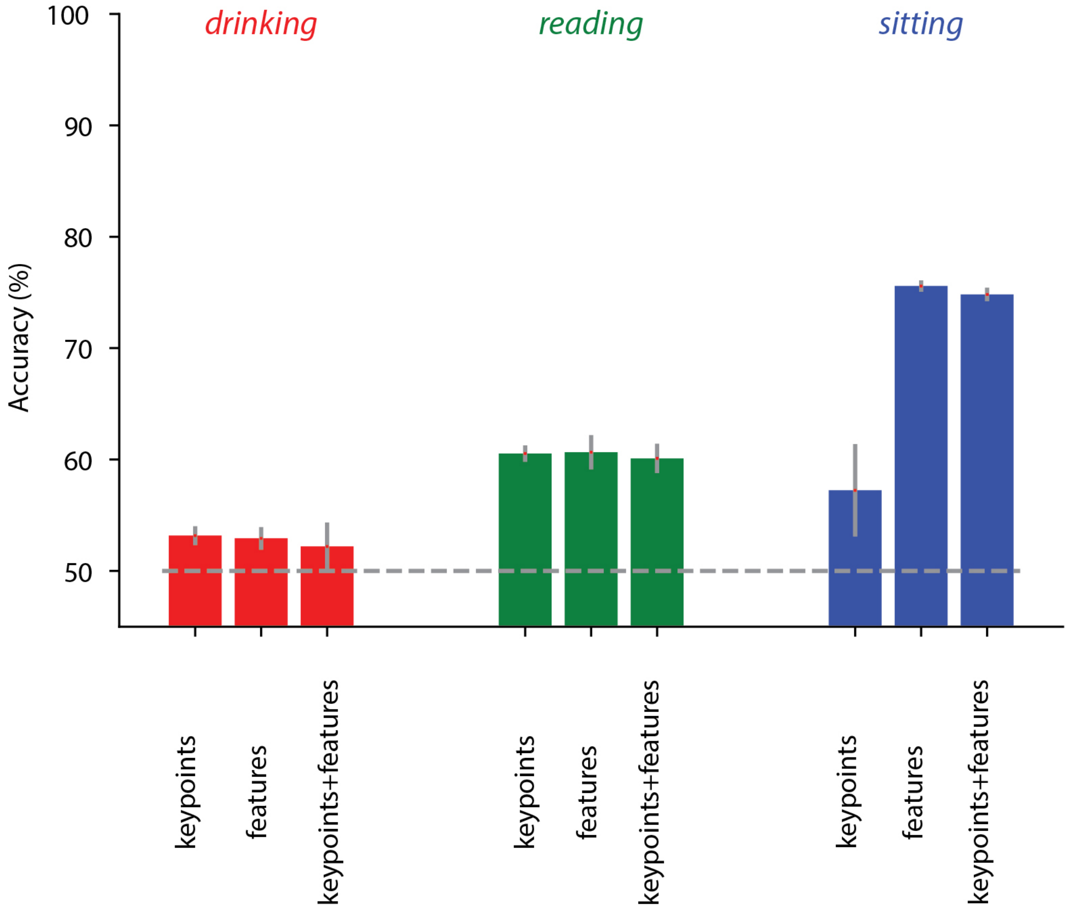

Figure S2. **Performance of detectron models extracting task-relevant features using RGB images**. This figure follows the conventions and format of **Figure 7** in the main text. Here we present results using RGB images. We extracted specific *keypoints* and *features* using the Detectron algorithm [1] (see main text for details).

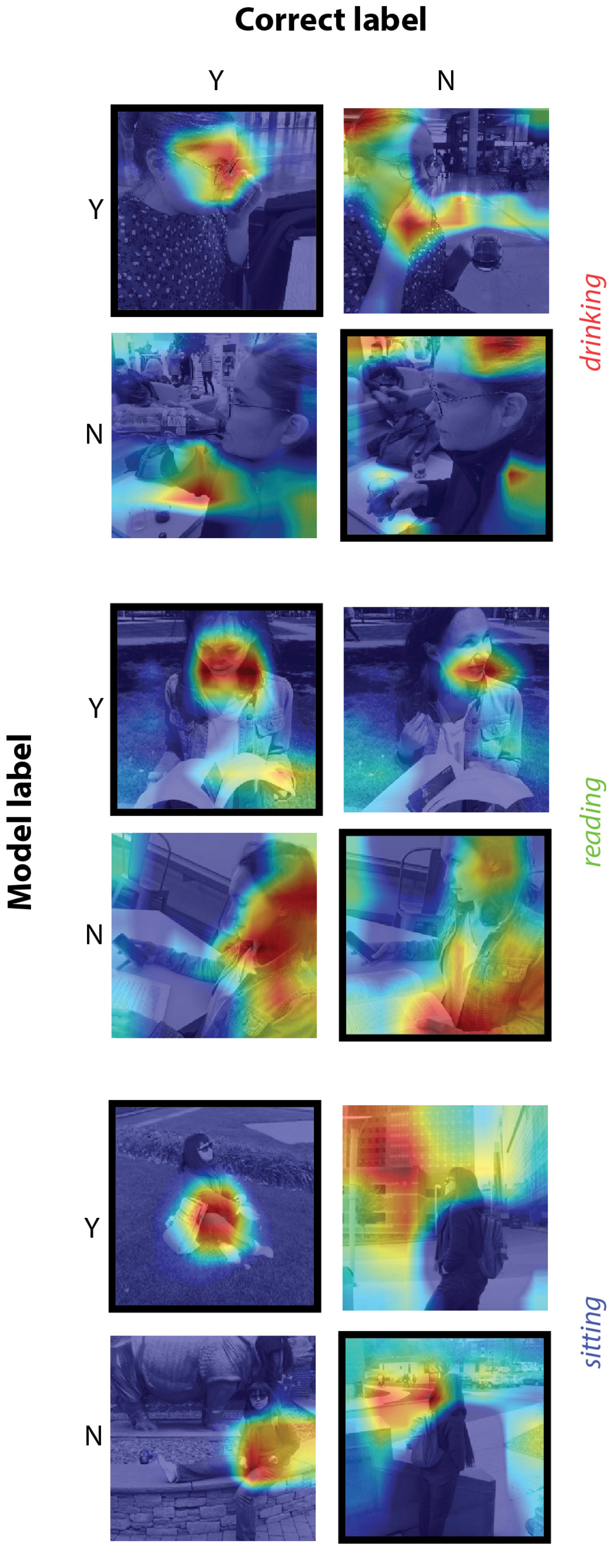

Figure S3. **Visualization of relevant features used by the network for classification**. Visualization of the salient features using Grad-CAM [3] for the ResNet-50 network [2] with weights pre-trained on ImageNet, finetuned on either the drinking, reading or sitting datasets. The gradient is used to compute how each feature contributes to the predicted class of a picture. On the last convolutional layer, the values of the features translate to a heatmap (red for most activated, blue for least activated). The heatmap is resized from 8x8 to 256x256 such as to overlap the input image.

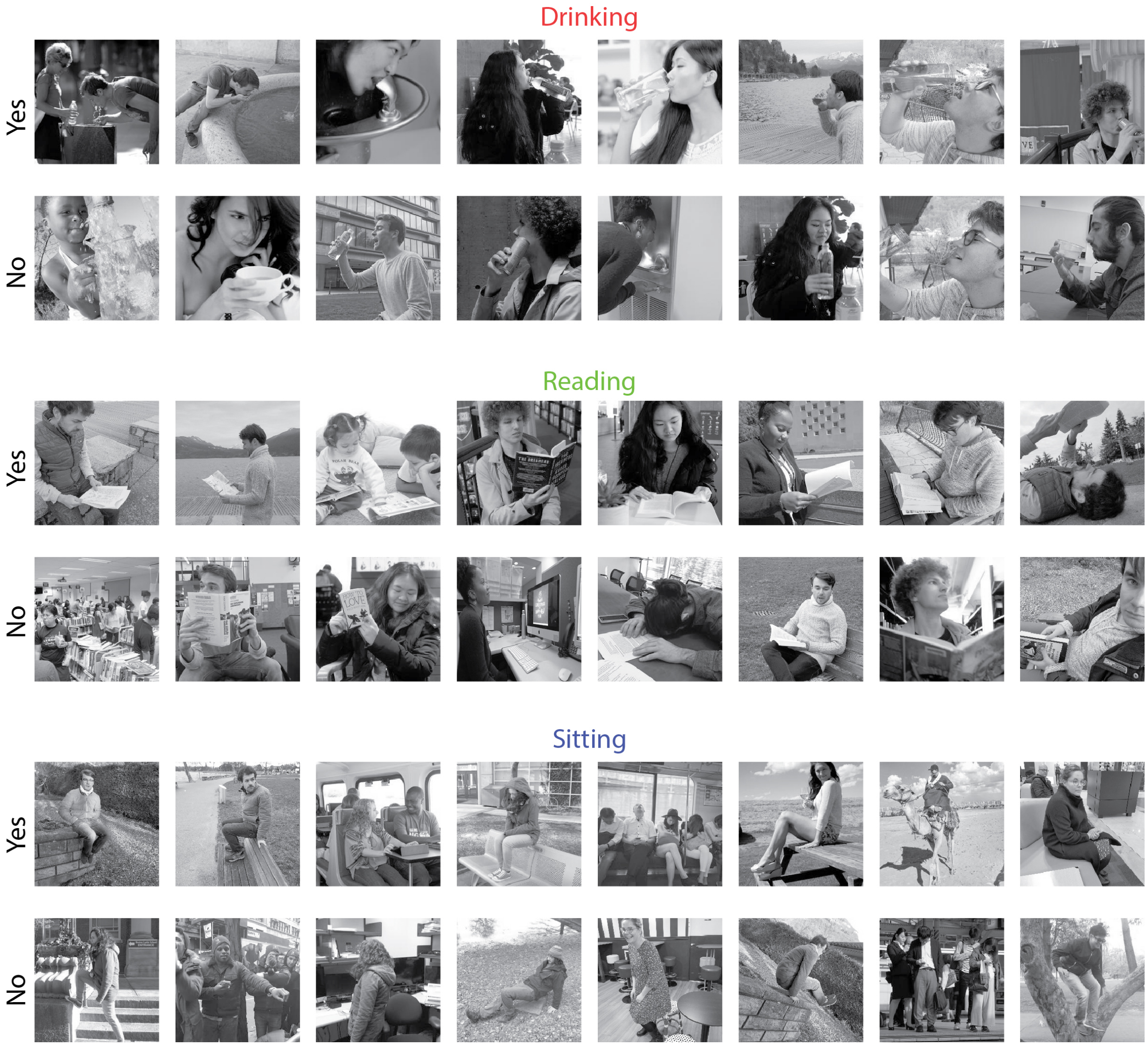

Figure S4. **Example images from our dataset.**

References

[1] Ross Girshick, Ilija Radosavovic, Georgia Gkioxari, Piotr Dollár, and Kaiming He. Detectron. https://github.com/facebookresearch/detectron, 2018. 2

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. *CoRR*, abs/1603.05027, 2016. 3

[3] Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. *CoRR*, abs/1610.02391, 2016. 3