

## **Description of Additional Supplementary Files**

**File name: Supplementary Movie 1**

**Description:** Video prediction results by comparison model trained with and without gradient masking.

**File name: Supplementary Movie 2**

**Description:** Video prediction results by SuperAnimal model fine-tuned with naive-fine-tuning and SuperAnimal model fine-tuned with memory replay.

**File name: Supplementary Movie 3**

**Description:** Video prediction results by SuperAnimal-TopViewMouse model with and without spatial pyramid inference. Note that because we use a detector for the SuperAnimal-Quadruped, this is not needed.

**File name: Supplementary Movie 4**

**Description:** Video prediction results by SuperAnimal models with and without video adaptation.

**File name: Supplementary Movie 5**

**Description:** Example video from Sturman et al (15) vs. SuperAnimal-TopViewMouse without any training.

**File name: Supplementary Movie 6**

**Description:** Top-down based SuperAnimal-TopviewMouse's video prediction from one example MABe video, without being trained on any MABe videos.