Test-Time Training is a general approach for improving the performance of predictive models when training and test data come from different distributions. It adapts to a new test distribution on the fly by optimizing a model for each test input using self-supervision before making the prediction. This method improves generalization on many real-world visual benchmarks for distribution shifts. In this talk, I will present the recent progress in the test-time training paradigm. I will show how masked auto-encoding overcomes the shortcomings of previously used self-supervised tasks and improves results by a large margin. In addition, I will demonstrate how test-time training extends to videos - instead of just testing each frame in temporal order, the model is first fine-tuned on the recent past before making a prediction and only then proceeding to the next frame.