GPPs FPGA accelerator Forward propagation Knowledge distilling Reference fixed-point model(teacher) Teacher network Softmax \mathbf{z}^{T} p^T(Soft target) CNN configuration 000 Softmax with loss Student network Softmax with loss $L(p^T, p^S)$ Input image data $L(y, p^S)$ $\mathbf{z}^{\mathbf{S}}$ 000 Initial fixed-point model(student) Student loss y(hard target) Backward propagation Label from input image data Student network Update weights