



Test plan

NetApp Solutions

NetApp
October 20, 2021

This PDF was generated from <https://docs.netapp.com/us-en/netapp-solutions/ai/ai-edge-test-plan.html> on October 21, 2021. Always check docs.netapp.com for the latest.

Table of Contents

Test plan 1

Test plan

[Previous: Technology overview.](#)

This document follows MLPerf Inference v0.7 [code](#), MLPerf Inference v1.1 [code](#), and [rules](#). We ran MLPerf benchmarks designed for inference at the edge as defined in the follow table.

Area	Task	Model	Dataset	QSL size	Quality	Multistream latency constraint
Vision	Image classification	Resnet50v1.5	ImageNet (224x224)	1024	99% of FP32	50ms
Vision	Object detection (large)	SSD-ResNet34	COCO (1200x1200)	64	99% of FP32	66ms
Vision	Object detection (small)	SSD-MobileNetsv1	COCO (300x300)	256	99% of FP32	50ms
Vision	Medical image segmentation	3D UNET	BraTS 2019 (224x224x160)	16	99% and 99.9% of FP32	n/a
Speech	Speech-to-text	RNNT	Librispeech dev-clean	2513	99% of FP32	n/a
Language	Language processing	BERT	SQuAD v1.1	10833	99% of FP32	n/a

The following table presents Edge benchmark scenarios.

Area	Task	Scenarios
Vision	Image classification	Single stream, offline, multistream
Vision	Object detection (large)	Single stream, offline, multistream
Vision	Object detection (small)	Single stream, offline, multistream
Vision	Medical image segmentation	Single stream, offline
Speech	Speech-to-text	Single stream, offline
Language	Language processing	Single stream, offline

We performed these benchmarks using the networked storage architecture developed in this validation and compared results to those from local runs on the edge servers previously submitted to MLPerf. The comparison is to determine how much impact the shared storage has on inference performance.

[Next: Test configuration.](#)

Copyright Information

Copyright © 2021 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means-graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system-without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark Information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.