Qualcomm Developer Project Mobile Multimodal Visual Language Model

Project Submission

|  |  |  |
| --- | --- | --- |
| **Attribution:** | Yunlong zhao | |
| **Email address** | <zhangzz6687@thundersoft.com>  [siyuan.he@thundersoft.com](mailto:hongliang.liu@thundersoft.com)  yunlong.zhao[@thundersoft.com](mailto:tianye.fan@thundersoft.com) | |
| **Project Title**\* | **Mobile Multimodal Visual Language Model** | |
| **Images**  *Upload up to 5 images of your project*  *Please submit/send the original JPEG/PNG files for all images included in the document* | **C6490P DK**    [Alt tag: “Blurred Image Clearness Processing using the C6490P Develop Kit”]  **Type-c usb line**   |  | | --- | | **typc** |   [Alt tag: “using the USB line to develop on C6490P Develop Kit” ]  **Charger**  charger  [Alt tag: “using round-hole charger to power C6490P Develop Kit”] | |
| **Description**\*  *High level description of the project* ***(75 words or less)*** | MobileVLM is a powerful multimodal visual language model (MMVLM) designed to run on mobile devices. It is the fusion of countless mobile-oriented architectural designs and technologies, including a set of 1.4B and 2.7B parameter scale language models trained from scratch, and a multi-modal vision model pre-trained on: via efficient projectors Perform CLIP stylish, cross-modal interactions. MobileVLM is evaluated against several typical VLM benchmarks. The model shows equivalent performance compared to some larger models. | |
| **Objective**   * *What inspired you to create this project?* * *What is your desired outcome?* | MobileVLM is a multi-modal visual language model (MMVLM). llama.cpp achieves high end-side performance through inference of Meta's LLaMA model (and other models) in pure C/C++. At the same time, it can be applied to the field of intelligent robots to achieve intelligent features. | |
| **Materials Required / Parts List / Tools** | Part Name | Link to purchase |
| C6490P DK | https://www.thundercomm.com/product/c6490p-development-kit/ |
| USB Line | https://item.jd.com/40759941966.html |
| Charger | https://www.thundercomm.com/product/c6490p-development-kit/ |
| **Source Code / Source Examples / Application Executable**  *Link to open source / shareable code repository* | Description | Link |
| Source Code | https://github.com/ThunderSoft-XA/MobileLLVM-on-6490 |
| **Additional Resources**  *List related links or resources such as websites, videos, presentations, or other materials* | Resource Title | Link or File Name (and provide file) |
| Video | https://github.com/ThunderSoft-XA/MobileLLVM-on-6490/doc/usage.mp4 |
| MobileVLM original model | <https://huggingface.co/liuhaotian/llava-v1.5-7b>  https://huggingface.co/openai/clip-vit-large-patch14-336 |
| MobileVLM modified model | https://github.com/ThunderSoft-XA/MobileLLVM-on-6490/cpp/model |
| Model porting tutorial | https://github.com/ThunderSoft-XA/MobileLLVM-on-6490/doc/MobileVLM-on-6490.doc |
| **Build / Assembly Instructions** | 1. cd cpp/source 2. mkdir build & cd build 3. cmake ../ 4. make -j6 5. make llava-cli 6. cp bin/llava-cli .. 7. cd .. 8. ./llava-cli -m ../model/ggml-model-q4\_k.gguf --mmproj ../model/mmproj-model-f16.gguf --image ../input/input.jpg -p 'what is it in the picture' | |
| **Usage Instructions** | The Demo running results are as follows：  final result: | |
| **Contributor(s) Info**  *Feel free to include headshots!* | Name | Title  Company |
| <zhangzz6687@thundersoft.com> | Thundersoft |
| [yunlong.zhao@thundersoft.com](mailto:tianye.fan@thundersoft.com) | Thundersoft |

––– Continued on next page –––

Filters and Tags for QDN projects page

|  |  |  |
| --- | --- | --- |
| **Platform/Hardware** | CSR 101x/102x Bluetooth  DragonBoard 410c  mangOH Red/Yellow  Qualcomm C8550 | MDM920x LTE for IoT  QCA-402x WiFi/BLE/Zigbee  Qualcomm Robotics RBx Dev Kit |
| **Software Tools** | 3D Audio Plugin for Unity  Adreno GPU SDK  Hexagon DSP SDK | Neural Processing SDK for AI  　Snapdragon Profiler |
| **Operating System** | Android  √ Linux  ThreadX RTOS | √ Ubuntu Core  Windows 10 IoT Core |
| **Cloud Services/Platform** | Sierra Wireless AirVantage  Gizwits Cloud Platform  AT&T M2X  IBM Bluemix | IBM Watson IoT  Microsoft Azure IoT  Amazon AWS IoT |
| **Skill Level Required** | Advanced  Beginner  √ Intermediate |  |
| **Areas of Focus** | 3D Printing & Modeling  Alexa Voice Service  √ Artificial Intelligence  Bluetooth  √ Computer Vision  Digital Signage  Education  √ Embedded  Gaming | Healthcare  IoT  √ Robotics  Security  Sensors  Smart Cities  Smart Home  Toys |

*By submitting your content (“Submission”), you are granting Qualcomm a royalty-free, perpetual, non-exclusive, unrestricted, worldwide license to: (a) post, use, copy, sublicense, adapt, transmit, publicly perform or display any such Submission, (b) use, reproduce, modify, adapt, publish, translate, create derivative works from, distribute, perform, play, host, communicate, make available and publish your Submission without restriction and (c) sublicense to third parties the unrestricted right to exercise any of the foregoing rights granted with respect to the Submission. The foregoing grants shall include the right to exploit any ideas, concepts, intellectual property, or proprietary rights in such Submission, including but not limited to rights under copyright, trademark, servicemark or patent laws under any relevant jurisdiction without Qualcomm owing any monies to you whatsoever. You represent and warrant that you own all right, title and interest in and to the Submission, or you have been granted sufficient rights in and to the Submission allowing the foregoing use of such Submission.*