



به نام خدا

Hardware for AI - بهار ۱۴۰۴



## تمرین چهارم : پیاده‌سازی سطح پایین شبکه‌های عصبی کانولوشنی

طراح: بهناز نکونام، امید صالحی نیا

### هدف پروژه :

هدف این تمرین آشنایی شما با نحوه استقرار مدل‌های یادگیری عمیق روی سخت افزارهای دارای منابع محدود مانند برد های توسعه سیستم های نهفته است.

### مقدمه:

در این تمرین شما باید مدل شبکه عصبی کانولوشنی را که در تمرین‌های قبلی آموزش داده و چندی سازی کرده اید را روی برد Z-turn اجرا و عملکرد آن را بسنجید. در این تمرین قصد داریم با استفاده از محیط توسعه نرم افزاری Vitis پروژه ای بسازیم، کد های مربوط به شبکه و ورودی ها را اضافه و مدل خود را روی برد اجرا کنیم.

### آماده سازی محیط و فایل ها:

- (1) آماده‌سازی محیط و فایل‌ها:  
ما در VIVADO، بلوکی طراحی میکنیم و فایل با پسوند XSA را در اختیار شما قرار می‌دهیم. دیتاست و وزن های مدل خود را به فرمتی مانند h. تبدیل کنید و در پوشه ای خیره کنید.
- (2) ایجاد پروژه نرم افزاری در Vitis :  
نرم افزار Vitis را باز کنید. یک پروژه جدید بسازید و فایل XSA که ما برای شما آماده کردیم را بعنوان سخت افزار انتخاب کنید. یک پروژه نرم افزاری (application project) بسازید که هدف ان اجرای مدل روی پردازنده ARM روی برد است.
- (3) اضافه کردن کد ها به پروژه :  
کد به زبان C و فایل های هدر h. را به پروژه اضافه کنید. اطمینان حاصل کنید که تنظیمات مسیرهای include و کتابخانه‌ها صحیح است.
- (4) ساخت و بارگذاری پروژه روی برد:

پروژه را build کنید تا کد c کامپایل شود و فایل اجرایی تولید شود. دیباگ و سپس اجرا کنید. نتایج را از طریق ترمینال سریال نرم افزار مشاهده کنید. خروجی را شامل accuracy و loss را گزارش دهید.

(5) زمان استنتاج را گزارش کنید.

(6) پیشتر در تمرین قبل، چندی سازی برای ۱۶، ۸ و ۶ بیت پیاده سازی کرده اید. اکنون مراحل ۱ تا ۵ را برای آن ها دوباره تکرار کنید.

(7) **(امتیازی)** در تمرین قبل با پروفایل کردن برنامه، بخش هایی که زمانبر هستند و می توانند به صورت موازی اجرا شوند مشخص شد. در این بخش می خواهیم با استفاده از شتابدهنده NEON پردازنده Cortex 9 این بخش ها را شتابدهی کنیم. در نظر داشته باشید که زمان اجرا باید با بخش های پیشین مقایسه شود. توضیحات تکمیلی این بخش :

NEON یک واحد پردازش SIMD (Single Instruction Multiple Data) در معماری ARM است که برای پردازش موازی داده ها طراحی شده. تراشه های Zynq شرکت Xilinx که شامل پردازنده ARM Cortex-A9 هستند، از NEON پشتیبانی می کنند. در این تمرین شما میتوانید از قابلیت این واحد پردازشی برای انجام عملیات های محاسباتی شبکه عصبی استفاده کنید. برای این کار ابتدا می بایست توابعی برای عملیات های مورد نیازتان ایجاد کرده و جایگزین تابع های قبلی خود کنید تا بتوانید به صورت موازی در هر مرحله داده های خود را پردازش کنید. برای استفاده از این واحد پردازش شما می بایست هدر arm\_neon.h را به کد های خود اضافه کرده و flag های کامپایلر را در قسمت

```
C/C++ build --> settings --> tool settings --> ARM v7 gcc --> Miscellaneous --> "-mfpu=vfpv3"
: "-mfpu=neon"
```

به neon تغییر دهید. در ضمن اگر از کتابخانه استاندارد استفاده کردید حتما آن را در قسمت

```
"ARM v7 gcc linker --> Libraries --> "-lm
```

اضافه کنید. برای مثال اگر از کتابخانه math استفاده کردید -lm را اضافه کنید.

به عنوان یک مثال تابع زیر ۸ عدد صحیح ۸ بیت را با ۸ عدد صحیح ۸ بیت دیگر به صورت موازی جمع کرده و خروجی را در آرایه ۸ تایی ۸ بیت قرار میدهد.

```
void add_u8_neon(const uint8_t *a, const uint8_t *b, uint8_t *resu)
```

```
{
```

```
    // Load 8-bit values into NEON 8-lane vector
```

```
    uint8x8_t va = vld1_u8(a); // Load 8 unsigned 8-bit integers
```

```
    uint8x8_t vb = vld1_u8(b);
```

```
    // Add the vectors
```

```
    uint8x8_t vsum = vadd_u8(va, vb);
```

```
// Store the result
vst1_u8(result, vsum);
}
```

در صورت وجود هرگونه اشکال میتوانید با ایمیل زیر در ارتباط باشد:

[mahdimn2011@yahoo.com](mailto:mahdimn2011@yahoo.com)

[behnaznekoona7@gmail.com](mailto:behnaznekoona7@gmail.com)

[omidn2@gmail.com](mailto:omidn2@gmail.com)

## سایر نکات

- \_\_\_\_ - انجام این تمرین به صورت گروه های دوفره خواهد بود.
- فایل ها و گزارش خود را با نام HWAI\_HW5\_<SID>.zip به ترتیب در محل های مربوطه در صفحه درس آپلود کنید.
- نام گذاری صحیح متغیرها، تمیزی کد و توضیحات می تواند تا حدودی کاستی های کد را در بخش های دیگر جبران کند.
- - هدف این تمرین یادگیری شماسست! در صورت کشف تقلب، مطابق با قوانین درس برخورد خواهد شد.
- به ازای هر x روز تأخیر، ۲ به توان x (با شمارش از صفر) از نمره شما کسر خواهد شد.

موفق باشید