دانشگاه شهید بهشتی

دانشکده مهندسی و علوم کامپیوتر

اطلاعیه دفاع

نام دانشجو: سیده سودابه موسوی	**نام استاد راهنما: دکتر دارا رحمتی** **دکتر قاسم جابری پور**	
مقطع: کارشناسی ارشد	**رشته: مهندسی کامپیوتر**	**گرایش: معماری سیستم‌های کامپیوتر**
نوع دفاع: • **دفاع پروپوزال** □ • **دفاع پایان نامه** ■ • **دفاع رساله دکترا** □	**تاریخ: ۱۴۰۳/۰۶/۲۴** **ساعت: ۱۵-۱۷** **مکان: کلاس ۱۱۷**	

عنوان: طراحی شتاب‌دهنده سخت‌افزاری شبکه‌های عصبی کارا با به‌کارگیری سیستم اعداد پیمانه‌ای

داوران خارجی: دکتر شاهین حسابی	**داوران داخلی: دکتر حمیدرضا مهدیانی، دکتر حسن عطارزاده**

چکیده: در سال‌های اخیر، شبکه‌های عصبی عمیق به‌طور گسترده‌ای در زمینه‌های متنوعی، از جمله پردازش تصویر پزشکی، سیستم‌های وسایل‌نقلیه خودران، بینایی کامپیوتری و تشخیص گفتار، و حتی تشخیص سرطان و دیگر کاربردها در سال‌های اخیر افزایش یافته‌است که بخشی از آن به‌دلیل عملکرد بی‌سابقه آن‌ها در حل مسائل مختلف به کار می‌رود. با این حال، بهبود دقت در این شبکه‌ها معمولاً به بهای افزایش پیچیدگی آنها تمام می‌شود. برای کاهش این پیچیدگی ضمن حفظ دقت، روش‌های قدرتمند و رایجی مانند کاهش تعداد پارامترها یا فعال‌سازی‌ها معرفی‌شده‌اند. با این وجود، مطالعات قبلی نشان داده‌اند که با افزایش طول عملوندهای ورودی، نمایش عدد دودویی استاندارد (BNS) قادر به پاسخ‌گویی به نیازهای سیستم‌های کارآمد از نظر مساحت، زمان و توان به‌دلیل مشکلات زنجیره انتشار نقلی نمی‌باشد. از این رو روش‌های نوآورانه‌ای مانند سیستم‌های عدد مانده‌ای (RNS) و سیستم‌های اعداد افزونه‌ای برای غلبه بر این چالش معرفی‌شده‌اند که کارایی محاسباتی را به‌طور قابل توجهی افزایش می‌دهند. یکی از ویژگی‌های اساسی سیستم‌های اعداد مانده‌ای، انتشار رقم نقلی محدود به درون پیمانه‌ها در عملیات حسابی است که باعث بهبود سرعت اجرای عملیات حسابی می‌شود. این انتشار محدود را می‌توان با نمایش رقم‌های مانده‌ای به‌صورت افزونه کاهش داد. معماری پیشنهادی به نام RRNS-Net، یک واحد شتاب‌دهنده مبتنی بر سیستم عددی مانده‌ای افزونه‌ای است که با بهره‌گیری از جریان داده در معماری (Eyeriss v2) طراحی شده‌است. این سیستم به‌طور مؤثر از محاسبات شبکه عصبی با حذف انتشار رقم نقلی در حین عملیات حسابی بین ارقام پشتیبانی می‌کند و در نتیجه اجرای سریع‌تری را فراهم می‌آورد. این معماری شامل پیاده‌سازی مدارهای محاسباتی و توابع فعال‌سازی ReLU در قالب R-RNS است. عملکرد معماری شتاب‌دهنده پیشنهادی با استفاده از شبکه‌های معروفی مانند AlexNet، VGG16 و SqueezeNet مورد ارزیابی قرار گرفته است. نتایج حاکی از آن است که RRNS-Net در مقایسه با پیاده‌سازی‌های دودویی و مانده‌ای به‌ترتیب به سرعت‌های محاسباتی ۱.۳۷ برابر و ۱.۰۵ برابر دست می‌یابد. علاوه بر این، مصرف انرژی در مقایسه با پیاده‌سازی دودویی در طول عملیات جمع و ضرب متوالی ۲۷٪ بهبود می‌یابد.