

اطلاعیه دفاع

نام دانشجو: فاطمه سپه وند		نام استاد راهنما: دکتر دارا رحمتی	
مقطع: کارشناسی ارشد		رشته: مهندسی کامپیوتر	
نوع دفاع:		گرایش: معماری سیستم های کامپیوتر	
<ul style="list-style-type: none"> <input type="checkbox"/> دفاع پروپوزال <input checked="" type="checkbox"/> دفاع پایان نامه <input type="checkbox"/> دفاع رساله دکترا 		تاریخ: ۱۴۰۳/۰۹/۱۰	
		ساعت: ۱۰:۳۰-۱۲:۰۰	
		مکان: کلاس ۳۰۱	
عنوان: شتاب‌دهی محاسبات شبکه‌های عصبی پیچشی با استفاده از محاسبات تصادفی			
داوران خارجی: دکتر شاهین حسابی		داوران داخلی: دکتر مهدیانی	
چکیده:			
<p>یک شبکه عصبی پیچشی معمولی از تعداد زیادی لایه تشکیل شده است که لایه‌های پیچشی بیشترین مقدار بار محاسبه و هزینه سخت‌افزار را تشکیل می‌دهند. در این شبکه‌ها مجموعه‌ای از نگاشت و ویژگی ورودی با مجموعه‌ای از فیلترها ترکیب شده است. هر نقطه در نگاشت و ویژگی خروجی با کشیدن فیلترها بر روی کانال‌های نگاشت و ویژگی ورودی مربوطه و محاسبه ضرب داخلی هر پنجره و سپس جمع‌بندی در تمام کانال‌ها به دست می‌آید. به دلیل تعداد زیاد ضرب در هر لایه، توسعه طرح‌های کم‌هزینه برای این عملیات سنگین اجتناب‌ناپذیر است. محاسبات تصادفی یک جایگزین کم‌هزینه برای پیاده‌سازی‌های دودویی مرسوم شبکه‌های عصبی پیچشی است. در این پژوهش، ما یک روش نوین برای افزایش سرعت محاسبات پیچشی در شبکه‌های عصبی پیچشی (CNN) از طریق کاهش تعداد پالس‌های ساعت و استفاده از محاسبات تصادفی و اشتراک گذاری وزن ارائه می‌دهیم. با اعمال الگوریتم‌های خوشه‌بندی بر روی وزن‌های تعداد مشخصی از کانال‌ها در ماتریس وزن لایه‌های پیچشی و تعیین تعداد بهینه خوشه‌ها، وزن‌ها را می‌توان به مراکز خوشه‌های مربوطه کاهش داد. این روش دقت شبکه‌های عصبی را حفظ کرده و در عین حال تکرار وزن‌ها را افزایش می‌دهد. با محاسبه تفاوت وزن‌ها، تعداد صفرهای ذخیره‌شده در بافر وزن بیشتر شده و در نتیجه تعداد پالس‌های ساعت کاهش می‌یابد.</p> <p>بنابراین با کمتر کردن تعداد پالس‌های ساعت، سرعت سخت‌افزار حداقل $1/43$ و $4/5$ برابر (با کاهش 18% و 17% مساحت با استفاده از تکنیک RLE) و حداکثر $4/27$ و $13/02$ برابر به ترتیب نسبت به معماری‌های SkippyNN و BISC-MVM در شبکه‌ی عصبی LeNet-5 افزایش داده می‌شود. همچنین در شبکه‌ی عصبی AlexNet حداقل $1/32$ و $1/67$ برابر (با کاهش 25% و 23% مساحت با استفاده از تکنیک RLE) به ترتیب نسبت به معماری‌های SkippyNN و BISC-MVM افزایش داده می‌شود.</p>			