对于 $\forall$ x, y $\in$ C 与 $\forall\, \lambda \in [0,1]$, 有 $\lambda x + (1-\lambda)y \in C$

由于全平面是一个凸集，故任何平面点集都可用全平面盖住，即能被凸集盖住，从而盖住该凸集的所有凸集的交集存在，即凸包存在.

而如果某个凸集 A 有两个凸包 M1 与 M2, 则 $M1 \bigcap M2$ 也能盖住凸集 A，且 $M1 \bigcap M2 \subset M1$，但 M1 是 A 的凸包，故 $M1 \subset M1 \bigcap M2$，故 $M1 \bigcap M2 = M1$. 同理 $M1 \bigcap M2 = M2$. 即 $M1 = M2$

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = x_1y_2 + x_3y_1 + x_2y_3 - x_3y_2 - x_2y_1 - x_1y_3$$

$\mathcal{O}(n \log n)$ $\mathcal{M}(n \log n)$ mx

It is worth noting that there is no simple linear relation between d and r with performances measured by mIoU, though d = 8r is a satisfactory choice. Experiments show that even r = 8 performs well, revealing that it can be very cheap for modeling the global context We test more optimization steps in the evaluation stage. In general, the same K for training and test is recommended. K = 6 is enough for CD and NMF, while even K = 1 is acceptable for VQ. Typically 3 6 steps are enough since simple MD' s prior is still biased, and full convergence can overfit it. The few iterations are cheap and act as early stopping. This section shows the advantages of MD-based Hamburger over attention-related context modules in computational cost, memory consumption, and inference time. We compare Hamburger (Ham) with self-attention (SA) (Vaswani et al., 2017), Dual Attention (DA) module from DANet (Fu et al., 2019), Double Attention module from A2 Net (Chen et al., 2018b), APC module from APCNet (He et al., 2019b), DM module from DMNet (He et al., 2019a), ACF module from CFNet (Zhang et al., 2019b), reporting parameters and costs of processing a tensor Z R 1×512×128×128 in Tab. 3. Excessive memory usage is the key bottleneck of cooperating with attention in real applications. Hence we also provide the GPU load and inference time on NVIDIA TITAN Xp. In general, Hamburger is light in computation and memory compared with attention-related global context modules.