

	White box attack	Black box attack
	<p>攻撃方法としては非現実的 最悪ケースの想定とモデル理解のため</p>	<p>現実的な攻撃方法 実際の攻撃シーンを想定</p>
<p>Digital attack</p> <p>画像ピクセル値を 直接操作して攻撃</p>	<p>モデルの脆弱性の理解： モデルが有する脆弱性の検証や攻撃 に対する最悪ケースの想定に利用</p> <p>例) [1610.08401] Universal adversarial perturbations</p>	<p>デジタルで稼働するシステムへ攻撃： API サービスへの攻撃（人をゴリラと 認識させ訴訟するなど）などに利用</p> <p>例) [1905.07121] Simple Black-box Adversarial Attacks</p>
<p>Physical attack</p> <p>被写体に操作をし て攻撃 （明るさ、距離、 角度などが変化する環境で攻撃）</p>	<p>対象物への摂動有効性の理解： 対象物に付与する摂動の諸条件下で の頑健性の検証や transferability を利 用した black box 攻撃に利用</p> <p>例) [1707.08945] Robust physical- world attacks ...</p>	<p>物理環境で稼働するシステムへ攻撃： 自動運転への攻撃（一時停止を速度 制限に認識させるなど）などに利用</p> <p>例) [1804.05810] ShapeShifter: ... （black box は transferability を利 用）</p>